



UNIVERSIDADE
E D U A R D O
MONDLANE

FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos factores associados a seropositividade em
crianças expostas ao HIV/SIDA no Distrito da Matola,
no primeiro semestre de 2022**

Autor: Alfredo Baúle Júnior

Maputo, 21 de Março de 2024



UNIVERSIDADE
E D U A R D O
MONDLANE

FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos factores associados a seropositividade em
crianças expostas ao HIV/SIDA no Distrito da
Matola, no primeiro semestre de 2022**

Autor: Alfredo Baúle Júnior

Supervisor: Miranda Albino Martins Muualo, PhD, UFRJ

Maputo, 21 de Março de 2024

Dedicatória

À minha mãe **Justina Tomas Nhatsave**

Pois sem ela não chegaria até aqui.

Que Deus continue lhe abençoando hoje, amanhã e sempre.

Declaração de Honra

Declaro por minha honra que o presente Trabalho de Licenciatura é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de Licenciado em Estatística, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, 21 de Março de 2024

(Alfredo Baúle Júnior)

Agradecimentos

O meu primeiro agradecimento é a **Deus**, pela proteção, pela saúde e pelas forças para que eu pudesse enfrentar todos os obstáculos nesta caminhada que foi concluída com muita alegria e satisfação.

Ao meu orientador, **Doutor Miranda Albino Martins Muualo**, pelo acompanhamento durante este trabalho, pela oportunidade deste aprendizado e por toda a valiosa contribuição nesta pesquisa.

Agradeço ao **Dr. Tozé José Namalela**, pelas horas e apoio disponibilizado em todas as fases da elaboração, organização assim como o final do prazo de entrega da pesquisa.

Agradeço também à todos meus colegas do Serviço Distrital de Saúde Mulher e Acção Social da Matola, pela disponibilidade e facilitação para obtenção da informação da pesquisa em particular ao colega de profissão **Almirdo Eugénio José Mangué**.

Aos colegas, amigos, **Castro Rone Edmundo Maússe, Valdemiro Eusébio, Nunes Lisboa, Chande Jamine, Neima Esperança, Hélder Guambe, Catarina Massango, Euflásio Francisco, Tereza Farinja, Alfredo Tiago, Clara Lopes, Cidália Nhambi, Reginaldo Julião, Nelson Basílio, Rivaldo Boa, Dionísia Rodrigues, Dionilza Júlio, Natália Lázaro e Estevão Zeca** pela paciência, incentivo, apoio durante estes quatro anos que estivemos juntos na faculdade.

À **Tânia Cossa**, pelo incentivo, confiança, dedicação, apoio para que o desenvolvimento desta pesquisa fosse uma realidade.

Aos colegas de Licenciatura em Estatística e colegas da Universidade pelas horas de estudo, por todas as suas contribuições nos trabalhos, debates e seminários. Também agradeço a todos que de uma forma ou de outra contribuíram para concretização deste trabalho.

Muito Obrigado!

Resumo

A seropositividade em crianças expostas ao HIV/SIDA é um factor que está directamente ligado às mulheres grávidas seropositivas, devido a não adesão as informações e as práticas de prevenção consequentemente aumentando o risco de transmissão vertical em recém-nascidos e aumentando desta maneira a taxa de positividade. Neste contexto, o presente estudo visa analisar os factores associados à seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola, no 1º semestre de 2022. Foi realizado um estudo retrospectivo, com uma abordagem quantitativa que contou com uma amostra de 300 observações, a recolha dos dados ocorreu nos dias 5 à 23 de dezembro de 2022 e o mesmo foi estendido ao dia 28 de dezembro de 2022. A análise estatística foi feita mediante a aplicação do teste Qui-quadrado de Pearson, teste de Wall, teste de hosmer-lemeshow, teste de razão de máxima verossimilhança e modelos de regressão logística múltipla. Relativamente ao perfil sócio-demográfico das crianças expostas ao HIV/SIDA, 45% eram do sexo masculino e 55% eram do sexo feminino. Em análise da associação entre as variáveis sociodemográficas, clínicas, sociais e económicas foi feita análise bivariada usando o teste qui-quadrado com todas as variáveis independentes em função da dependente onde constatou-se que as variáveis que demonstraram ter alguma associação foram as seguintes: medicamentos, alimentação, início tardio, abandono, religião, carga viral e pTV da criança. Portanto as mesmas variáveis foram analisadas usando a regressão logística onde observou-se as seguintes variáveis: Medicamentos, carga viral, abandono, religião e alimentação. No que respeita à razão de chances a variável que aponta as dificuldades da mãe em se alimentar teve a seguinte classificação ($OR= 8.95$; IC 95% 1.74; 45.86), quanto a toma regular dos medicamentos em ($OR= 9.042$; IC 95% 2.49; 32.82, carga viral não suprimida ($OR=3.74$; IC 95% 1.23; 11.31), nenhuma carga em ($OR=7.09$; IC 1.70; 29.56) e por motivo de abandono ($OR=44.35$; IC 3.20; 613.85). O estudo mostrou ainda que, há uma necessidade de melhorar as estratégias assim como as políticas de saúde no acesso a qualidade dos cuidados na consulta de criança em risco (CCR) e consulta pré-natal (CPN), com objectivo de reduzir a taxa de seropositividade nas crianças expostas ao HIV.

Palavras-Chave: Seropositividade em crianças expostas; Factores de Risco; Modelos de Regressão Logística.

Abstract

Seropositivity in children exposed to HIV/AIDS is a factor that is directly linked to HIV-positive pregnant women, due to non-adherence to information and prevention practices, consequently increasing the risk of vertical transmission in newborns and thus increasing the rate of positivity prevention. In this context, the present study aims to analyze the factors associated with seropositivity in children exposed to HIV/AIDS in the District of Matola, in the first half of 2022. A retrospective study was carried out, with a quantitative approach that involves a sample of 300 observations, data collection occurred from December 5th to 23rd, 2022 and the same was extended to December 28th, 2022. Statistical analysis was carried out through the application of Pearson's Chi-square test, Wall's test, Hosmer's test - lemeshow, maximum likelihood ratio test and multiple logistic regression models. Regarding the sociodemographic profile of children exposed to HIV/AIDS, 45% were male and 55% were female. In analyzing the association between sociodemographic, clinical, social and economic variables, and bivariate analysis was carried out using the chi-square test with all independent variables depending on the dependent one, where it was found that the variables that demonstrated some association were the following: medications, nutrition, late ART initiation, abandonment, viral load, religion and child's PTV. Therefore, the same variables were analyzed using logistic regression where the following variables were observed: medications, nutrition, viral load, abandonment and religion. With regard to the odds ratio, the variable that points out the mother's difficulties in feeding had the following classification (OR= 8.95; 95% CI 1.74; 45.86), as for taking medications regularly (OR= 9.042; 95% CI 2.49; 32.82, unsuppressed viral load (OR=3.74; 95% CI 1.23; 11.31), no viral load (OR=7.09; CI 1.70; 29.56) and due to dropout (OR=44.35; CI 3.20; 613.85). The study also showed that, there is a need to improve strategies and health policies in access to quality of care in child at risk consultation (CCR) and prenatal consultation (CPN), with the aim of reducing the rate of seropositivity in children exposed to HIV.

Keywords: Seropositivity in exposed children; Risk factors; Logistic Regression Models.

Abreviaturas

AUC Area Under the ROC Curve

CCR Consulta de Criança em Risco

CNCS Conselho Nacional de Combate ao HIV/SIDA

COB Centro de Investigação Operacional da Beira

CPN Consulta Pré-Natal

DTS Doença de Transmissão Sexual

FHI360 The Science of Improving Lives

HIV Vírus da Imunodeficiência Humana

INS Instituto Nacional de Saúde

INSIDA Inquérito Nacional sobre o Impacto do HIV e SIDA em Moçambique

MISAU Ministério da Saúde

SIDA Síndrome de Imunodeficiência Adquirida

SMI Serviço Materno Infantil

TARV Tratamento Antirretroviral

TMI Transmissão Materna Infantil

TV Transmissão Vertical

UNAIDS Joint United Nations Programme on HIV/SIDA (ONUSIDA)

UNC Carolina Population Center

US Unidade Sanitária

ONG organização não governamental

OMS Organização Mundial de Saúde

Índice

Dedicatória.....	i
Declaração de Honra	ii
Agradecimentos	iii
Resumo	iv
Abstract.....	v
Abreviaturas.....	vi
Lista de Figuras	x
Lista de Tabelas	xi
Introdução.....	1
1.1. Contextualização.....	1
1.2. Definição do problema.....	2
1.3. Objectivos	2
1.3.1. Objectivo Geral.....	2
1.3.2. Objectivos Específicos	2
1.4. Relevância do Estudo	3
1.5. Estrutura do Trabalho	3
Revisão de Literatura.....	5
2.1. Conceitos Básicos	5
2.2. História do HIV	6
2.2.1. Situação do HIV no Mundo e na África.....	7
2.2.2. Situação do HIV em Moçambique e na Província de Maputo	8
2.2.3. Factores de Risco Relacionados com Infecção do HIV	8
2.3. Modelos lineares generalizados.....	9
2.3.1. Características do modelo linear generalizado.....	10
2.4. Análise de regressão	11

2.4.1. Regressão logística.....	11
2.4.2. Modelo de regressão logística binária.....	12
2.2.3. Regressão logística múltipla.....	12
2.2.4. <i>Odds</i> e linearização da função logística	14
2.2.5. Interpretação dos coeficientes de regressão do logit	15
2.2.6. Estimação dos coeficientes da regressão	18
2.2.7. Associação entre as variáveis do estudo.....	19
2.2.7.1. Teste de <i>qui-quadrado de Pearson</i> χ^2	19
2.2.8. Testes de significância sobre os coeficientes do logit	20
2.2.8.1 <i>Pseudo R²</i>	20
2.2.8.2. Teste de razão de verossimilhança	21
2.2.8.3. Teste de Wald	22
2.2.9. Teste de ajustamento do modelo	23
2.2.9.1. Teste de Hosmer-Lemeshow	23
2.2.10. Outros modos de avaliar a qualidade do modelo de regressão logística	24
2.2.10.1. Poder de classificação (sensibilidade e especificidade)	24
2.2.10.2. Curva <i>ROC (Receiver Operating Characteristic)</i>	25
2.2.11. Diagnóstico de <i>outliers</i> e de observações influentes	26
2.2.12. Verificação dos pressupostos do modelo logístico.....	27
2.2.13. Seleção do melhor modelo	28
2.2.14. Estudos Relacionados	30
Material e Métodos.....	32
3.1 Material.....	32
3.1.1 Local do estudo.....	32
3.1.2. Classificação do estudo	32
3.1.3. População, amostra e amostragem	34
3.1.4. Critérios de Inclusão e Exclusão	35

3.1.4.1. Inclusão.....	35
3.1.4.2. Exclusão	35
3.1.5. Variáveis de estudo.....	36
3.2. Métodos	37
3.2.1. Técnicas da estatística usadas (procedimentos do estudo).....	37
CAPÍTULO 4	39
Resultados e Discussão.....	39
4.1. Apresentação dos resultados.....	39
4.1.1 Descrição do perfil da amostra	39
4.1.2 Perfil sociodemográfico e económico das crianças expostas ao HIV/SIDA no Distrito da Matola.....	39
4.1.3 Análise da associação das variáveis do estudo.....	43
4.1.4 Análise de regressão logística e estimativa de modelo de regressão logística para os factores associados a seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola.....	44
4.2. Discussão dos resultados	51
Conclusões e Recomendações	53
5.2. Recomendações	54
5.3. Limitações	54
Referências Bibliográficas.....	55
Anexos.....	59

Lista de Figuras

Figura 4.1: Distribuição das crianças expostas por sexo.....	39
Figura 4.2: Análise de outliers do modelo2.....	45
Figura 4.3: <i>Receiver Operating Characteristic Curve (ROC)</i> do modelo de regressão logística para a determinação da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola.	48

Lista de Tabelas

Tabela 2.2.3: Classificação do modelo.....	24
Tabela 2.2.4: Valor indicativos da área <i>ROC</i>	25
Tabela 3.1: Referente as unidades sanitárias onde foi efectuado a recolha dos dados.....	32
Tabela 3.2: Aplicação da amostragem probabilística estratificada para a determinação de amostra por cada unidade sanitária.....	35
Tabela 3.3: Descrição das variáveis do estudo.....	36
Tabela 4.1: Perfil sociodemográfico e económico das crianças expostas ao HIV/SIDA no Distrito da Matola.....	40
Tabela 4.2: Perfil social das crianças expostas ao HIV/SIDA no Distrito da Matola.....	41
Tabela 4.3: Perfil clínico da mãe e das crianças expostas ao HIV/SIDA no Distrito da Matola.....	42
Tabela 4.4: Intervalo de confiança a 95% de Odds Ratio para categorias de variáveis significativas.....	44
Tabela 4.5: Comparação dos modelos construídos através do AIC e BIC.....	45
Tabela 4.6: Análise de multicolinearidade do modelo2.....	46
Tabela 4.7: Teste de razão de verossimilhança para avaliação global do modelo.....	46
Tabela 4.8: Modelo de regressão logística para a determinação dos factores associados a seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola.....	47
Tabela 4.9: Teste de Hosmer e Lemeshow.....	48
Tabela 4.10: Matriz de cruzada do modelo de regressão logística a determinação da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola.....	49
Tabela 4.11: Modelo logístico para para a determinação da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola	49

CAPÍTULO 1

Introdução

1.1. Contextualização

O vírus da Imunodeficiência Humana (HIV) é o agente causador da Síndrome de Imunodeficiência adquirida (SIDA), que é uma doença infecciosa, crônica e fatal que leva a destruição do sistema imunológico (Pereira, 2013).

Segundo à WHO (2023), o HIV continua a ser um importante problema de saúde pública global, configurando-se num desafio aos diversos sectores sociais, quanto às medidas de controle das suscetibilidades ligados aos aspectos individuais e contextuais de exposição ao vírus aonde as estimativas apontam que no final de 2022, havia cerca de 39,0 milhões [33,1–45,7 milhões] de pessoas que viviam com HIV e dois terços das quais (25,6 milhões) vivem na Região Africana da organização mundial de saúde (OMS). Contudo em 2022, 630 000 [480 000–880 000] pessoas morreram de causas relacionadas com o HIV e 1,3 milhões [1,0–1,7 milhões] de pessoas contraíram o HIV.

Moçambique situa-se entre os oito países com maior taxa de prevalência estimada de HIV em adultos na idade produtiva. Dados da Roda de Vigilância de 2009 mostram que a prevalência de HIV de adultos 15-49 anos em Moçambique é de 15% duas vezes superior à média subsariana de 7,2% e estima-se que 45% das novas infecções ocorrem no grupo etário dos 15-24 anos (Insida, 2009).

De acordo com Onusida (2015), entende-se como criança exposta ao HIV, aquela nascida de mãe infectada ou que tenha sido amamentada por mulheres infectadas pelo HIV. A Saúde Materno Infantil é considerada um grupo alvo prioritário nos cuidados e tratamento em saúde, visto que o maior índice de vulnerabilidade ao desenvolvimento das crianças esta virado para mulheres. Portanto foi criada a consulta da criança em risco onde são encaminhadas todas crianças dos 0 aos 18 meses, com exposição ao HIV, assim como crianças com um cuidado específico, com desnutrição crônica, exposição a Tuberculose entre outros casos.

Contudo no presente trabalho de pesquisa pretende-se avaliar os factores associados a seropositividade em crianças expostas ao HIV/SIDA na Província de Maputo em Moçambique, visto que o tema é de extrema importância olhando aquilo qui é o aumento das novas infecções no Distrito da Matola no 1º semestre de 2022 em 2% comparativamente a todo ano de 2021 que apresentou 3.2% do total de crianças infectadas pelo vírus do HIV/ SIDA.

1.2. Definição do problema

A maior parte das mães actualmente são jovens e adolescentes que se encontram ainda na idade escolar, por vezes sem condições para sustentabilidade dos seus filhos recém-nascido sem contar também com a falta de apoio social e económico por parte da família. Apesar de altos investimentos feitos no PTV (Prevenção de Transmissão Vertical), ainda residem casos positivos de crianças expostas ao HIV, a prova disso são os constantes debates que tem se levantado em quase todas as reuniões de cruzamento e discussão de dados mensais no Distrito da Matola, visto que existem muitas ONG's (Organizações Não Governamentais) que trabalham em coordenação com o MISAU (Ministério da Saúde) com objectivo de reduzir a contaminação das crianças expostas ao HIV. Por outro lado as mães tem um tratamento assegurado por parte das US's (Unidades Sanitárias) do Distrito da Matola. Entretanto segue a seguinte questão:

- Quais são os factores associados a ocorrência de casos positivos em crianças expostas ao vírus de HIV no Distrito de Matola?

1.3. Objectivos

1.3.1. Objectivo Geral

Analisar os factores associados a seropositividade em crianças expostas ao HIV no Distrito da Matola no 1º semestre de 2022.

1.3.2. Objectivos Específicos

- Descrever o perfil sócio-demográfico das crianças expostas ao vírus HIV no Distrito da Matola;
- Avaliar associação entre a variável de interesse com as variáveis explicativas;
- Estimar o modelo de Regressão Logística relacionada a seropositividade das crianças expostas ao HIV no Distrito da Matola.
- Identificar os factores que influenciam na infecção por HIV nas crianças expostas ao vírus no Distrito da Matola;

1.4. Relevância do Estudo

Actualmente muitas adolescentes e jovens tornam-se mães em contexto de total desconhecimento dos cuidados a ter com a sua vida sexual, bem como sob influência acentuada do comportamento dos mesmos que levam dia-pós-dia. Atenção as gestantes, puérperas e crianças no sentido de prevenir a morbimortalidade materna e infantil é uma prioridade que aponta para necessidade de adoptar medidas destinadas a assegurar a melhoria do acesso da cobertura da qualidade do acompanhamento pré-natal, assistência ao parto e da criança (Calvo, 2021). A Saúde Materno Infantil (SMI), representa um sector prioritário a nível nacional assim como de todo mundo, onde engloba a consulta da criança e risco (CCR) que tem como objectivo o atendimento de todas as crianças com prematuridade, baixo peso, desnutrição, exposição ao HIV entre outros casos.

Dado que o HIV é um problema de Saúde Pública que tem merecido atenção do Governo e parceiros de implementação, mesmo assim tem se verificado um número elevado de crianças expostas ao HIV, nesse âmbito realizar o estudo do gênero é de extrema importância porque vai fornecer aos tomadores ou gestores de informação sobre quais são os factores associados a seropositividade em crianças expostas ao HIV, além de que esse estudo é aplicado a uma região específica do Distrito da Matola, sendo que a nível da literatura pouco se sabe sobre estudos do gênero, o que sugere que evidências científicas sejam necessárias.

No âmbito académico o estudo trás o fundamento que é possível aplicar a teoria para situações de respostas de problemas de Saúde Pública.

Na comunidade é relevante nos resultados que o estudo vai trazer, no sentido de despertar atenção da mesma para que possam mudar de atitudes.

1.5. Estrutura do Trabalho

O presente trabalho é constituído por cinco capítulos, descrito como:

- Capítulo 1 - Introdução: apresenta a contextualização do tema, a definição do problema de estudo, a questão de pesquisa, os objectivos do estudo, a relevância e a estrutura do trabalho;
- Capítulo 2 - Revisão de Literatura: apresenta conceitos básicos, abordagens do HIV no mundo, Africa, Moçambique e Província de Maputo, Regressão Binária, Modelos Lineares Generalizados especificamente o Modelo de Regressão Logística Múltipla, bem como alguns estudos similares;

- Capítulo 3 - Material e Métodos: neste capítulo, são apresentados de forma detalhada os materiais usados e métodos aplicados para a produção dos resultados que são apresentados na secção 4 desta pesquisa
- Capítulo 4 - Resultados e Discussão: Apresenta os resultados obtidos da análise estatística dos dados e discussão dos mesmos de forma comparativa com alguns estudos disponíveis.
- Capítulo 5 - Conclusão e Recomendações: Apresenta as conclusões obtidas do trabalho, suas limitações e as recomendações.

Ainda também, consta na parte final as referências utilizadas no desenvolvimento do trabalho e materiais suplementares – Anexos.

CAPÍTULO 2

Revisão de Literatura

2.1. Conceitos Básicos

Segundo Feracin (2002), o **vírus da imunodeficiência humana (HIV¹)** é um retrovírus que pertence à família Retroviridae da subfamília Lentivirinae, que ataca as células do sistema imunitário (o sistema de defesa do organismo humano), fazendo com que pouco a pouco, essas células tornam-se incapazes de proteger o organismo humano contra infecções e tumores. Razão pela qual as pessoas que têm SIDA apanham doenças que o seu organismo não consegue combater.

Segundo Seidl et al. (2005), o **síndrome de imunodeficiência adquirida (SIDA)** é uma doença causada por um vírus chamado HIV. Esta doença, ainda não existe cura e nem uma vacina porém actualmente existem medicamentos capazes de controlar a infecção e permitir uma vida mais saudável e longa. As letras da palavra SIDA significam o seguinte:

S - Síndrome: conjunto de sinais e sintomas;

I - Imuno: sistema imunitário, o sistema de defesa do organismo humano;

D - Deficiência: não funciona bem;

A - Adquirida: foi contraída e não é hereditária.

Formas de Transmissão do HIV

Sexual - representa a principal forma de exposição em todo o mundo, sendo que a transmissão heterossexual, nas relações sem o uso de preservativo é considerada pela OMS (Organização Mundial de Saúde) como a mais frequente.

Sanguínea (em receptores de sangue e em usuários de drogas injectáveis) - a transmissão sanguínea associada ao uso de drogas injectáveis é um meio muito eficaz de transmissão do HIV, devido ao uso compartilhado de seringas e agulhas.

Vertical (da mãe para o filho, durante a gestação, parto ou por aleitamento) - a transmissão vertical, decorrente da exposição da criança durante a gestação, parto ou aleitamento materno, vem aumentando devido à maior transmissão heterossexual.

¹ Human Immuno Deficiency Virus

Além dessas formas, mais frequentes, também pode ocorrer a transmissão ocupacional, ocasionada por acidente de trabalho, em profissionais da área da saúde que sofrem ferimentos com instrumentos perfuro-cortantes contaminados com sangue de pacientes infectados pelo HIV.

Prevenção do tratamento vertical (PTV)

Segundo Vaz et al. (2011), a transmissão vertical mãe-filho através da placenta e durante o parto ou gestação nas mães infectadas, representa quase a totalidade dos casos de SIDA nas crianças expostas ao HIV. O PTV surge como medidas ou estratégias educacionais de prevenção contra a transmissão vertical (TV) do HIV de mãe para filho.

Criança exposta ao HIV

Segundo o Misau (2018), criança exposta ao HIV é aquela que durante a gestação ou parto, tenha sido nascida e amamentada por mulher infectada pelo HIV.

Consulta da criança em risco (CCR)

É um sector que está contido no serviço materno infantil (SMI) que admite crianças nas seguintes condições: prematuridade (nascer antes do tempo), baixo peso ao nascer (inferior a 2,5 kg), Gémeos, desmame brusco em menores de 1 ano de idade, aleitamento artificial em menores de 1 ano de idade, malnutrição, crianças com atraso no desenvolvimento, doenças graves, filhos de mães HIV-positivas, todas as crianças com mais de 3 internamentos em 12 meses, contacto com tuberculose pulmonar, infecções respiratórias, entre outras doenças Cncs (2009). Portanto as crianças expostas ao HIV são tratadas na CCR.

2.2. História do HIV

Segundo Feracini (2002), a Síndrome da Imunodeficiência Adquirida (SIDA) foi reconhecida em meados de 1981, nos EUA, a partir da identificação de um número elevado de pacientes adultos do sexo masculino, homossexuais e moradores de São Francisco ou Nova York, que apresentavam sarcoma de Kaposi, pneumonia por *Pneumocystis carinii* e comprometimento do sistema imune, o que levou à conclusão de que se tratava de uma nova doença, ainda não classificada, de etiologia provavelmente infecciosa e transmissível.

De acordo com Vaz et al. (2011), o primeiro caso de SIDA em Moçambique foi diagnosticado em 1986 e até Março de 1996 o Ministério da Saúde (MISAU) já notificara cerca de 4,906 casos à Organização Mundial da Saúde (OMS). O Programa Nacional de Combate às Doenças de

Transmissão Sexual/SIDA (PNC/ITS/SIDA) do MISAU estima que o número de casos até agora notificados corresponda somente a 5% do total dos casos de SIDA que os serviços de saúde deveriam diagnosticar e notificar, se não pesassem os constrangimentos de natureza técnica e logísticos (acesso). Na verdade, a fraca cobertura sanitária, a falta de pessoal qualificado e de meios de diagnóstico, dentre outros aspectos, dificultam a notificação duma proporção considerável de casos de SIDA que ocorrem em Moçambique.

2.2.1. Situação do HIV no Mundo e na África

Segundo Unids, (2009), há cerca de 33.4 milhões de pessoas que vivem com HIV, mostrando que a epidemia continua a crescer desde 1981 quando foi pela primeira vez reconhecida, fazendo dela a epidemia mais destrutiva da história recente. A maior parte das pessoas infectadas encontra-se nos primeiros anos de vida economicamente activa e são responsáveis de famílias. Esforços sustentados em diversas regiões do mundo têm ajudado a reduzir a incidência da infecção por HIV entre os diferentes grupos de risco.

De acordo com Salgueiro et al. (2014), afirma que em todo o mundo são muitas as crianças afectadas pelo HIV, seja porque elas próprias são soropositivas ou porque têm algum parente próximo nessa condição. Estima-se que mais de um milhão de crianças em todo o mundo sejam portadoras do HIV e cerca de quatro em cada dez crianças HIV positivas morrem antes de completar um ano de idade.

Segundo a Onusida (2012), o continente África continua sendo o mais afectado pela epidemia do vírus, sendo responsável por 62% das novas infecções à nível mundial, contribuindo desta maneira em 69% das pessoas vivendo com o vírus. A escala da epidemia do HIV/SIDA varia em África assim como a sua seroprevalência, onde os países do norte e do nordeste da África tem taxas de prevalência de infectados significativamente mais baixos em 0.1% em relação a África Austral que é a região mais afectada no continente representando mais de 15% das novas infecções.

2.2.2. Situação do HIV em Moçambique e na Província de Maputo

De acordo com Ciob et al. (2018) Moçambique continua a enfrentar uma epidemia generalizada de HIV, que esta associada a um sistema de saúde ainda com deficiências tem afectado negativamente o desenvolvimento humano. Apesar das intervenções de prevenção do HIV nas últimas 3 décadas, o HIV/SIDA continua a ser um grande desafio de saúde pública em Moçambique. Portanto a meta de redução de novas infecções por HIV, melhorar o acesso aos cuidados e tratamento a todas as pessoas que vivem com o vírus, tem priorizado intervenções para as populações de maior risco para reduzir a incidência do HIV. Contudo, o sistema de saúde fraco, factores sociais e culturais, a deficiente coordenação entre os parceiros de implementação e o fraco uso de evidências na gestão de programas constituem algumas das questões desafiadoras que resultam em ineficiência na utilização dos recursos disponíveis para intervenções de HIV.

Segundo o InSida (2021), neste momento está bem claro que a prevalência do HIV em Moçambique está a aumentar cada vez mais em todo o país e principalmente nas províncias de Maputo Província e Cidade com prevalência do HIV em (15,4%) e (16,2%), Gaza com (20,9%), Sofala com (13,2%) e Zambézia com (17,1%) respectivamente.

A província de Maputo localiza-se na zona sul do país com uma área de 26058 Km² com uma população estimada em 2216460 milhões de habitantes segundo o censo populacional de 2017. De acordo com o InSida (2021) a prevalência do HIV na província é de (15,4%), tendo em consideração que a maioria das novas infecções tem ocorrido no grupo etário dos 15-24 anos. Contudo o Distrito da Matola vem registando aumentos significativos de novas infecções das crianças nascidas por mães seropositivas em situações preocupantes visto que todas as unidades sanitárias garantem a cobertura e seguimento de todas as mães positivas.

2.2.3. Factores de Risco Relacionados com Infecção do HIV

Segundo o Misau (2018), os factores associados ao aumento do risco de transmissão do HIV são:

- **Virais:** carga viral elevada e taxas baixas de circuncisão masculina;
- **Maternos:** estado clínico e imunológico comprometido (avançado), presença de ITS e outras co-infecções e estado nutricional deficiente;
- **Comportamentais:** reinfeção, seja por reexposição sexual ou compartilhamento de seringas, pouca ou nenhuma utilização do preservativo, grande proporção da população adulta com

múltiplos parceiros, vastas redes sexuais, “mistura de idades” geralmente entre homens mais velhos e mulheres jovens ou adolescentes;

- **Obstétricos:** trabalho de parto prolongado e tempo de ruptura das membranas amnióticas, via de parto, presença de hemorragia intraparto e parto instrumentalizado;
- **Inerentes ao recém-nascido:** prematuridade, baixo peso ao nascer e tempo de aleitamento materno.
- **Culturais:** visitas ao curandeiro, factores culturas de submissão ao homem e questões de tradição entre populações/culturas.

2.3. Modelos lineares generalizados

Segundo Favero et al., (2017), o estudo das distribuições estatísticas não é recente, e desde o início do século XIX, até aproximadamente o início do século XX, os modelos lineares que envolvem a distribuição normal praticamente dominou o cenário da modelagem de dados.

Entretanto, a partir do período entre guerras, começam a surgir modelos para fazer frente a situações em que as modelagens lineares normais não se adequavam satisfatoriamente. McCullagh e Nelder (1989), Turkman e Silva (2000) e Cordeiro e Demétrio (2007) citam, neste contexto, os trabalhos de Berkson (1944), Dyke e Patterson (1952) e Rasch (1960) sobre os modelos logísticos envolvendo as distribuições de Bernoulli e binomial, de Birch (1963) sobre os modelos para dados de contagem envolvendo a distribuição Poisson, de Feigl e Zelen (1965), Zippin e Armitage (1966) e Glasser (1967) sobre os modelos exponenciais, e de Nelder (1966) sobre modelos polinomiais envolvendo a distribuição Gama.

Todos estes modelos acabaram por ser consolidados, do ponto de vista teórico e conceitual por meio do seminal trabalho de Nelder e Wedderburn (1972), em que foram definidos os **Modelos Lineares Generalizados (Generalized Linear Models)**, que representam um grupo de modelos de regressão lineares e exponenciais não lineares, em que a variável dependente possui, por exemplo, distribuição normal, Bernoulli, binomial, Poisson ou Poisson-Gama.

Turkman e Silva (2000), os MLGs correspondem à uma síntese de modelos, vindo assim unificar, tanto do ponto de vista teórico como conceptual, à teoria da modelação estatística até então desenvolvida.

São, pois, casos particulares dos MLGs, os seguintes modelos:

- Modelo de regressão linear clássico;
- Modelos de análise de variância e covariância;
- Modelo de regressão logística;

- Modelo de regressão de Poisson.

2.3.1. Características do modelo linear generalizado

Segundo Fávero et al., (2009), os MLGs são caracterizados pela seguinte forma:

1. Componente aleatória: Dado o vetor de covariáveis \mathbf{x}_i das variáveis aleatórias Y_i de tamanho n , são condicionalmente idênticas e independentes com distribuição pertencente à família exponencial na forma canónica e, portanto, o seu valor médio é dado por:

$$f(Y_i, \mathbf{x}_i) = \mu_i = b'(\theta_i), i = 1, 2, \dots, n \quad (2.3)$$

2. Componente sistemática: Define-se o preditor linear η_i como combinação linear das variáveis explicativas, ou seja, a componente sistemática é formada pelas variáveis explicativas, que entram sob forma de uma soma linear de seus efeitos, dando origem a um vector de preditores lineares, representado pela expressão dado por:

$$\eta_i = \sum_{r=1}^r x_{ir} \beta_r = \mathbf{x}_i^T \boldsymbol{\beta} \text{ ou } \boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta} \quad (2.4)$$

$\mathbf{X} = (x_1, \dots, x_n)$ representa a matriz do modelo, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)$ o vector de parâmetros desconhecidos e $\boldsymbol{\eta}$ representa o vector de preditores lineares.

3. Função de ligação: Representa uma função que relaciona a componente aleatório ao componente sistemático, relacionando a média ao preditor linear, isto, é:

$$\eta_i = g(\mu_i) \quad (2.5)$$

Sendo $g(.)$ uma função monótona e diferenciável.

Contudo, para o presente estudo irá se fazer a referência do modelo de regressão logística.

2.4. Análise de regressão

Segundo Mestre (2009), o termo análise de regressão define um conjunto vasto de técnicas estatísticas usadas para modelar relações entre variáveis e predizer o valor de uma ou mais variáveis dependentes (ou de resposta) a partir de um conjunto de variáveis independentes (ou explicativas).

Existem várias metodologias estatísticas que permitem explicar ou descrever a relação entre uma variável de interesse (resposta) e uma ou mais variáveis (explicativas). O modelo linear é talvez o mais utilizado para modelar esta relação. Este modelo assume, entre outras, que o valor esperado da variável resposta é uma combinação linear das variáveis explicativas e que a variável resposta segue a distribuição normal.

O objectivo principal da análise de regressão é de estudar a relação entre variáveis, ou seja, analisar a influência que uma ou mais variáveis (explicativas), medidas em indivíduos ou objectos, têm sobre uma variável de interesse a que se dá o nome de variável resposta.

2.4.1. Regressão logística

Os modelos de regressão logística, embora bastante úteis e de fácil aplicação, ainda são pouco utilizados em muitas áreas do conhecimento humano. Embora o desenvolvimento de softwares e o incremento da capacidade de processamento dos computadores tenham propiciado a sua aplicação de forma mais directa, muitos pesquisadores ainda desconhecem as suas utilidades e, sobretudo, as condições para que seu uso seja correcto (Maroco, 2007).

Definição

De acordo com Gonzalez (2018), regressão logística é uma técnica estatística que tem como objectivo produzir a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária de uma ou mais variáveis independentes contínuas ou binárias.

Segundo Fávero et al. (2009), a regressão logística é uma técnica estatística utilizada para descrever o comportamento entre uma variável dependente binária e um conjunto de variáveis independentes métricas ou não métricas. Ou melhor, destina-se a investigar o efeito das variáveis pelas quais os objectos ou indivíduos estão expostos sobre a probabilidade de ocorrência de determinado evento de interesse exigindo que o resultado da análise possibilite associações a certas categorias, tais como sucesso ou insucesso.

De acordo com Gujarati e Porter (2011), a regressão logística possui uma variável dependente de carácter não métrica que é inserida através de uso de variáveis dummies, que tomam "0" para indicar a ausência de um atributo e "1" para indicar a presença do tal atributo.

Com base nas três definições acima citadas, o objectivo fundamental da regressão logística é prever uma variável dependente dicotómica a partir de um conjunto de variáveis independentes, mediante um modelo que apresenta um bom ajuste.

2.4.2. Modelo de regressão logística binária

Segundo Gonzalez (2018), a regressão logística binária ou univariada representa os casos de regressão logística em que a variável (Z) é binária ou dicotómica, ou seja, tem duas categorias e tem apenas uma variável independente.

Favero et al. (2017), afirma que a regressão logística binária tem como objectivo principal estudar a probabilidade de ocorrência de um evento definido por (Y) que se representa na forma dicotómica ($Y = 1$ para descrever a ocorrência do evento de interesse e $Y = 0$ para descrever a ocorrência do não evento), com base no comportamento de variáveis explicativas.

Seja uma variável resposta (Y) dicotómica, que assume valor (1) para evento de interesse (sucesso) e o valor (0) para evento complementar (fracasso), com probabilidade $\pi_i = P(Y = 1/X = x_i)$ e $1 - \pi_i = P(Y = 0/X = x_i)$, respectivamente. Para descrever a média condicional de (Y) dado (X) com a distribuição logística, usar-se-á a notação $\pi(x) = \pi_i = E(Y/x)$. O modelo de regressão logística simples pode ser representado pela equação (2.6).

$$y = \pi(x) + \varepsilon \quad (2.6)$$

Onde ε pode assumir um ou dois valores possíveis. Se $y = 1$ então $\varepsilon = 1 - \pi(x)$ com probabilidade $\pi(x)$ e se $y = 0$ então $\varepsilon = -\pi(x)$ com probabilidade $1 - \pi(x)$. Assim, ε tem uma distribuição com média zero e variância igual a $\pi(x)[1 - \pi(x)]$. Ou seja, a distribuição condicional da variável resposta segue uma distribuição binomial com probabilidade dada pela média condicional, $\pi(x)$ (Gonzalez, 2018).

2.2.3. Regressão logística múltipla

Segundo Figueira (2006), a regressão logística múltipla representa o contexto da regressão logística, em que, a variável dependente Y é binária ou dicotómica, ou seja, tem duas categorias e que há mais

de uma variável independente. Por exemplo temos uma variável dependente a ocorrência da morte ou não de um animal, e a variável independente é a dose aplicada para eutanásia, se adicionarmos uma variável independente como o peso animal, este caso deixa de ser regressão logística binária e passa a ser regressão logística múltiplo. Entretanto podemos considerar então a regressão logística múltipla como uma generalização da regressão logística binária.

Seja dado um conjunto de variáveis independentes, onde consideramos este conjunto denotado por

$X = (x_1, x_2, \dots, x_t)$. A combinação linear para este conjunto de variáveis é definida como:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_t x_t \quad (2.7)$$

Logo, o logit quando igualado à $g(x)$, é descrito na equação (2.9):

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_t x_t \quad (2.8)$$

Assim como na regressão logística binária, no caso múltiplo, utiliza-se o antilogaritmo na equação (2.9), para fins de isolar p obtendo assim, o modelo de regressão logística múltipla para a probabilidade estimada \hat{p} :

$$\hat{p} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_t x_t}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_t x_t}} = \frac{1}{1 + e^{-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_t x_t}} \quad (2.9)$$

A função de verossimilhança é a mesma da equação (2.10), com a diferença de que $\pi(x_i)$ é dado como $\pi(i)$, em função da equação (2.30) representado o conjunto de variáveis independentes em $g(x)$ e seus respectivos coeficientes. Logo a função log-verossimilhança é obtida como:

$$l(\beta) = \sum_{i=1}^n [y_i \ln \pi_i + (1 - y_i) \ln(1 - \pi_i)] \quad (2.10)$$

As expressões das equações a partir das derivadas parciais e são definidas pelas equações (2.11) e (2.12):

$$\frac{\partial l(\beta)}{\partial \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\pi}_i = 0 \quad (2.11)$$

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n x_{ij}y_i - \sum_{i=1}^n x_{ij}\hat{\pi}_i = 0, \text{ para } j \in \{1, \dots, t\} \quad (2.12)$$

$\hat{\pi}_i$ é o estimador de π_i , pelo método da máxima verossimilhança

2.2.4. Odds e linearização da função logística

Segundo Mangas (2019), define-se odds (chances ou rácio de verossimilhanças) por:

$$Odds = \frac{\pi}{1 - \pi} \quad (2.13)$$

Que traduz a razão entre as probabilidades do sucesso e de insucesso, isto é, a possibilidade de se observar o sucesso, $\{Y=1\}$, relativamente à possibilidade de se observar o insucesso, $\{Y=0\}$. Defina-se ainda a função que é o logaritmo neperiano do odds:

$$\text{logit}(\pi) = \ln\left(\frac{\pi}{1 - \pi}\right) \quad (2.14)$$

Conhecido por transformação logit. Se assumirmos que esta função é linear:

$$\text{logit}(\pi) = \beta_0 + \beta_1 x \quad (2.15)$$

Se tomarmos uma única variável independente (covariável) X , onde a constante β_0 é a ordenada na origem β_1 o coeficiente de regressão linear (declive), neste caso da função *logit*, ou

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (2.16)$$

Se tomarmos várias variáveis independentes x_1, x_2, \dots, x_k (k covariáveis), onde, novamente, β_0 é a constante e β_1, \dots, β_k são os coeficientes de regressão linear múltipla da função *logit*, facilmente se deduz que esta transformação *logit* lineariza à função logística:

$$\begin{aligned} \ln\left(\frac{\pi}{1-\pi}\right) &= \beta_0 + \beta_1 x \\ \Leftrightarrow \frac{\pi}{1-\pi} &= e^{\beta_0 + \beta_1 x} \\ \Leftrightarrow \pi(1 + e^{\beta_0 + \beta_1 x}) &= e^{\beta_0 + \beta_1 x} \end{aligned}$$

$$\Leftrightarrow \pi = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

$$\Leftrightarrow \pi = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (2.17)$$

ou, de modo similar,

$$\pi = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (2.18)$$

A transformação *logit* oferece assim maiores facilidades do ponto de vista matemático, mas também do ponto de vista da interpretação do próprio modelo.

2.2.5. Interpretação dos coeficientes de regressão do logit

De acordo com Gonzalez (2018), a regressão logística segue a distribuição de Bernoulli, portanto é preciso conectar as variáveis independentes à distribuição Bernoulli presente na variável dependente e esse *Link* é chamado de Logit. Na regressão logística nós não conhecemos a probabilidade p como é o padrão nos problemas de distribuição de Bernoulli, portanto o objectivo do modelo logístico é estimar p para uma combinação linear das variáveis independentes, onde o p estimado é probabilidade \hat{p} .

Para ligar a combinação linear de variáveis a distribuição de Bernoulli é necessária uma função que as una, ou mapeie a combinação linear de variáveis que poderiam retornar qualquer valor em uma distribuição de probabilidades de Bernoulli com um domínio de 0 a 1.

Segundo Hosmer e Lemeshow (2000), no modelo regressão linear simples assume-se que $E(Y | x) = \beta_0 + \beta_1 x$, em que $E(Y | x)$ é o valor esperado de Y dado o valor de x . Implica que é possível para $E(Y | x)$ assumir qualquer valor, como x varia entre $-\infty$ e $+\infty$.

Na regressão logística, assume-se que $\pi(x) = E(Y | x)$ representa a média condicional de Y dado x quando a distribuição logística é usada. A forma específica do modelo de regressão logística é (2.15).

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (2.19)$$

A transformação de $\pi(x)$ que é a base no estudo de regressão logística é a transformação logit. Essa transformação é definida, em termos de $\pi(x)$, como (2.16)

$$Z(x) = \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta_1 x \quad (2.20)$$

A razão de chances (*Odds Ratio* (OR)) é definida como a razão entre a chance de um evento ocorrer em um grupo e a chance de ocorrer em outro grupo, ou seja, trata-se que é uma medida de associação entre variáveis e pode ser aplicada em diversas áreas. Seu logaritmo natural, o logit, é esta função representada na equação:

$$\ln(Odds) = \ln \left(\frac{\pi}{1-\pi} \right) \quad (2.21)$$

Vamos considerar um modelo de regressão logística, o qual as variáveis independentes também são dicotômicas, isto é, a variável x será codificada em 0 ou 1. O modelo tem dois valores de $\pi(x)$ e equivalentemente dois valores de $1 - \pi(x)$.

Quando a variável independente X assume o valor $x = 1$ teremos:

$$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \quad e \quad 1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}} \quad (2.22)$$

E para $x = 0$, temos:

$$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \quad e \quad 1 - \pi(0) = \frac{1}{1 + e^{\beta_0}} \quad (2.23)$$

Isso quando a variável resposta Y assume os valores $y = 1$ e $y = 0$.

A chance da resposta quando $x = 1$ é definido como $\frac{\pi(1)}{1-\pi(1)}$ similarmente a chance da resposta quando $x = 0$ é $\frac{\pi(0)}{1-\pi(0)}$. O log da chance é dado por (2.24).

$$Z(1) = \ln \left[\frac{\pi(1)}{1-\pi(1)} \right] \quad e \quad Z(0) = \ln \left[\frac{\pi(0)}{1-\pi(0)} \right] \quad (2.24)$$

A razão das chances (*Odds Ratio*), é definida pela equação (2.13).

$$OR = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \quad (2.25)$$

O log da razão de chances denotado (*log-odds*), é a diferença do logit (2.14).

$$\ln(OR) = \ln \left[\frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \right] = Z(1) - Z(0) \quad (2.26)$$

Usando a expressão para o modelo de regressão logística mostrado nas equações (2.25) e (2.26), a razão de chances é definida por (2.27).

$$OR = \frac{\left[\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right] \left[\frac{1}{1 + e^{\beta_0}} \right]}{\left[\frac{e^{\beta_0}}{1 + e^{\beta_0}} \right] \left[\frac{1}{1 + e^{\beta_0 + \beta_1}} \right]} \quad (2.27)$$

E o log da razão de chances é dado por (2.28).

$$\ln(OR) = \ln(e^{\beta_1}) = \beta_1 \quad (2.28)$$

Deste modo, β_1 pode ser interpretado como sendo o logaritmo da razão de chances em condições acima referenciadas. Consequentemente, a razão das chances do sucesso versus insucesso quando a variável independente quantitativa x aumenta uma unidade versus manter-se constante é a exponencial do respetivo coeficiente de regressão do *logit*.

Convém recordar que a medida de efeito *OR* é um número real estritamente positivo e:

Se $OR < 1$, significa que as chances do evento de interesse (categoria em numerador) diminui em relação ao grupo que se está a analisar;

Se $OR = 1$, significa que não há associação entre ambas categorias ou que as duas categorias têm as mesmas chances de ocorrer;

Se $OR > 1$, significa que as chances do evento de interesse (da categoria em numerador) aumenta em relação ao grupo que se está a analisar.

2.2.6. Estimação dos coeficientes da regressão

De acordo com Gonzalez (2018), para ajustar um modelo de regressão é necessário estimar os parâmetros β_0 e β_1 do modelo. Para isso utiliza-se o método de estimação da máxima verossimilhança, a partir dos dados da amostra, ou seja, o conjunto de observações, este método irá procurar os estimadores para o modelo de regressão logística, que são os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ que maximizam o logaritmo da função de máxima verossimilhança. Portanto a estimação por máxima verossimilhança, permite encontrar os estimadores dos parâmetros do modelo que tem maior probabilidade de replicar o padrão de observações nos dados da amostra.

Seja $\beta = (\beta_0, \beta_1)$ o vector de coeficientes e sejam as probabilidades $P(Y_i = 1/x_i) = \pi(x_i)$ e $P(Y_i = 0/x_i) = 1 - \pi(x_i)$. Então para os pares (x_i, y_i) , tais que $y_i = 1$, a contribuição para a função de verossimilhança é $\pi(x_i)$, e para os pares tais que $y_i = 0$, a contribuição para a função de verossimilhança é $1 - \pi(x_i)$, onde $\pi(x_i)$ denota o valor de $\pi(x)$ avaliado em x_i .

A função de verossimilhança é dada pela equação (2.10):

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.29)$$

Aplicando-se o logaritmo natural em ambos os lados da equação obtemos a função log-verossimilhança:

$$l(\beta) = \ln[L(\beta)] = \sum_{i=1}^n [y_i \ln \pi(x_i) + (1 - y_i) \ln(1 - \pi(x_i))] \quad (2.30)$$

O valor β que maximiza $\ln[L(\beta)]$ é obtida após derivar $L(\beta)$ em relação os parâmetros (β_0, β_1) :

$$\frac{\partial \ln [L(\beta)]}{\partial \beta_0} = \sum_{i=1}^n [y_i - \pi(x_i)] \quad (2.31)$$

$$\frac{\partial \ln [L(\beta)]}{\partial \beta_1} = \sum_{i=1}^n x_i [y_i - \pi(x_i)] \quad (2.32)$$

Os estimadores de (β_0, β_1) , denotados por $(\hat{\beta}_0, \hat{\beta}_1)$, são as soluções das equações (2.30) e (2.31) quando iguados a 0. Estes estimadores dos parâmetros, medem a taxa do logit para uma unidade de variação na variável independente, isto significa que eles são de facto, a inclinação da linha de regressão entre a variável dependente y_i e a sua variável independente x_i .

Portanto as fórmulas matemáticas permitem aos programas de regressão logística identificar mais eficazmente, os estimadores que maximizam a função log-verossimilhança. Este algoritmo escolhe, sucessivamente, novos conjuntos de parâmetros que produzam maiores log-verossimilhança e melhores ajustamentos aos dados observados, isto é o processo continua através de ciclos repetitivos até maximização da função log-verossimilhança (Gonzalez, 2018).

O modelo de regressão logística fica então da seguinte forma:

$$\hat{\pi} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k}} = \frac{1}{1 + e^{-\left(e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k}\right)}} \quad (2.32)$$

2.2.7. Associação entre as variáveis do estudo

2.2.7.1. Teste de *chi-quadrado de Pearson* (χ^2)

Segundo Marôco (2011), o teste de chi-Quadrado de pearson (χ^2) serve para testar se dois ou mais grupos independentes diferem relativamente a uma determinada característica, isto é, se a frequência com que os elementos da amostra se repartem pelas classes de uma variável nominal categorizada é ou não idêntica. Para aplicação do teste é usada a seguinte estatística:

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{(L-1)(C-1)}^2 \quad (2.33)$$

Onde: O_{ij} - representa valores observados na linha "i" da coluna "j";

E_{ij} - representa valores esperados, na linha "i" da coluna "j";

L - é o número de linhas;

C - é o número de colunas.

Para obter a frequência esperada E_{ij} em cada célula, multiplicam-se os totais marginais comuns a uma determinada célula e divide-se pelo total de casos.

Hipóteses a testar:

H_0 : As variáveis são independentes.

H_1 : As variáveis são dependentes.

Rejeita-se a hipótese nula ao nível de significância se $\alpha \chi^2 > \chi^2_{(L-1)(C-1)}$ ou se $p - \text{value} < \alpha$.

Pressupostos do teste chi-quadrado de pearson

- Se o número de classes $k=2$, a frequência esperada mínima deveria ser maior ou igual a 5;
- Se $k > 2$, mais de 80% das frequências deveria estar acima de 5 e nenhuma delas com valor inferior a 1.

Logo, se um desses pressupostos fossem violados, foi usado o teste exacto de Fisher para verificar a associação (Petrie & Watson, 2013).

2.2.8. Testes de significância sobre os coeficientes do logit

2.2.8.1 *Pseudo R²*

Segundo o Gonzalez (2018), a estatística R^2 é uma medida que avalia em termos percentuais a qualidade de um ajustamento de um modelo de regressão linear aos dados observados. Por outra R^2 indica a percentagem de variância do modelo, ou seja, quanto de variação da variável dependente, está sendo explicada pelas variáveis independentes do modelo.

Por se tratar de regressão logística, não existe uma estatística que seja equivalente ao R^2 . A denominação de *pseudo-R²* deve-se ao facto de eles se parecem com um R^2 do modelo de regressão linear, pois estão em uma escala similar, variando de 0 à 1, mas apesar dessa similaridade não podem ser interpretados da mesma forma como se interpreta em regressão linear, visto que no modelo de regressão linear as variáveis dependentes são contínuas e na regressão logística a variável dependente é categórica (Batista, 2015).

Entretanto há muitas maneiras diferentes de calcular o *pseudo-R²*, porém neste trabalho, foi usado apenas o *pseudo-R²* "Nagelkerke", por ser um dos mais frequentemente usados pelos softwares estatísticos. Segue a sua definição na equação (2.34):

$$R_N^2 = \frac{R_{CS}^2}{1 - e^{-\frac{2LL_0}{n}}} \quad (2.34)$$

Onde:

n : representa o tamanho de amostra;

R_{CS}^2 : é o valor verossimilhança do modelo sendo estimado;

$2LL_0$: é o valor da função verossimilhança para um modelo sem preditores.

2.2.8.2. Teste de razão de verossimilhança

Uma vez estimado o modelo, é necessário testar a sua significância. Portanto para testar essa significância é usado o teste da razão de verossimilhança. Este teste avalia a significância dos coeficientes estimados simultaneamente, onde verifica se o modelo estimado é globalmente significativo, como também se os coeficientes de regressão associados β são todos nulos com exceção de β_0 (Cabral, 2013).

As hipóteses a testar são:

H_0 : $\beta_1 = \beta_2 = \dots = \beta_p = 0$ (isto é, o modelo não é estatisticamente significativo),

H_1 : $\exists_i: \beta_i \neq 0$ ($i = 1 \dots p$) (isto é, o modelo é estatisticamente significativo).

A estatística D tem como objectivo, comparar o modelo em análise e o modelo saturado, ou seja, a comparação dos valores observados e dos valores esperados usando a função de verossimilhança que é baseada na seguinte expressão:

$$D = -2 \ln \left[\frac{\text{verossimilhança de modelo ajustado}}{\text{verossimilhança de modelo saturado}} \right] \quad (2.35)$$

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (2.36)$$

Onde: modelo ajustado corresponde ao modelo com apenas as variáveis desejadas para o estudo e o modelo saturado corresponde ao modelo com todas as variáveis e interações. A função D também é chamada de *deviance* (desvio), sempre é positiva e quanto menor, melhor é o ajuste do modelo.

Segundo o Gonzavel (2018), para estimar a significância de uma variável independente, comparam-se o valor de D com e sem variável independente na equação. A alteração no valor de D esperado pela inclusão da variável independente no modelo é obtida através de:

$$G = D(\text{modelo sem } p \text{ variaveis}) - D(\text{modelo com } p \text{ variaveis})$$

$$G^2 = -2 \ln \left[\frac{L(R)}{L(F)} \right] \quad (2.37)$$

$L(R)$ a função de verossimilhança do modelo ajustado, isto é, o modelo contendo q parâmetros e $L(F)$ a verossimilhança do modelo completo.

G^2 segue aproximadamente uma distribuição χ^2_{p-q} , sendo a sua regra de decisão procedida de forma análoga ao teste G.

2.2.8.3. Teste de Wald

Segundo Silva (2011), o teste de Wald é também utilizado na regressão logística para a determinação da significância dos coeficientes do modelo estimado, onde testa se cada coeficiente é significativamente diferente de zero. Deste modo, o teste de Wald verifica se uma determinada variável independente possui uma relação estatisticamente significativa com a variável dependente. O teste de Wald é obtido comparando a estimativa de máxima verossimilhança de um coeficiente e a estimativa do seu erro padrão:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

A estatística de teste é dada por:

$$W_j = \frac{\hat{\beta}_j}{EP(\hat{\beta}_j)} \cap N(0,1) \quad (2.38)$$

Onde: $\hat{\beta}_j$ é o vector de estimativas para β , $EP(\hat{\beta}_j)$ é erro padrão de β_j e W_j é o teste que seguiu uma distribuição normal padrão.

2.2.9. Teste de ajustamento do modelo

Gonzalez (2018) afirma que após a estimação dos coeficientes, tem-se interesse em verificar a significância das variáveis no modelo assim como a qualidade de ajuste do modelo. Isto geralmente envolve formulação e teste de uma hipótese estatística para determinar se a variável independente no modelo é significativamente relacionada com a variável dependente como também se o modelo está bem ajustado.

2.2.9.1. Teste de Hosmer-Lemeshow

De acordo com Favero et al. (2009), O teste de Hosmer-Lemeshow é muito utilizado em regressão logística com a finalidade de testar a qualidade de ajuste do modelo, isto é, o teste avalia se o modelo obtido pode explicar adequadamente os dados observados. Este teste baseia-se na divisão das observações em grupos separados de dados de acordo com as probabilidades previstas. A estatística de teste é dada por:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - e_k)^2}{e_k \left(1 - \frac{e_k}{n_k}\right)} \quad (2.39)$$

Onde: $o_k = \sum_{j=1}^{n_k} y_{kj}$ representa o número de casos registados no k-ésimo decil e $e_k = \sum_{j=1}^{n_k} \hat{\pi}_{kj}$ representa o número esperado de casos no k-ésimo decil.

As hipóteses à testar são:

H_0 : O modelo logístico ajusta-se aos dados.

H_1 : O modelo logístico não se ajusta aos dados.

Sob H_0 , o modelo estimado ajusta-se aos dados, isto é, não existem diferenças significativas entre os valores previstos e os observados, em que a regra de decisão é:

- Se $\hat{C} \leq x_{1-x;g-2}^2$ ou $P(\hat{C} > x_{1-x;g-2}^2) \geq \alpha$, não rejeitar H_0
- Se $\hat{C} > x_{1-x;g-2}^2$ ou $P(\hat{C} > x_{1-x;g-2}^2) < \alpha$, rejeita H_0

2.2.10. Outros modos de avaliar a qualidade do modelo de regressão logística

2.2.10.1. Poder de classificação (sensibilidade e especificidade)

De acordo com Mangas (2019), a eficiência classificativa do modelo logístico pode também ser avaliada pela sensibilidade e especificidade apresentada pelo modelo.

Tabela 2.2.3: Classificação do modelo

		Valor Observado	
		Y = 1	Y = 0
Valor Estimado	$\hat{Y} = 1$	VP	FP
	$\hat{Y} = 0$	FN	VN

Onde: Verdadeiro Positivo (VP), Falso Positivo (FP), Falso Negativo (FN) e Verdadeiro Negativo (VN), P representa o total de “eventos” positivos (Y=1 ou “Resultado do teste positivo”) e N é o total de “não eventos” (Y=0, ou negativo, ou “Resultado do teste negativo”).

Classificação: representa à percentagem das predições correctas do modelo sobre o total.

$$AC = \frac{VP}{P + N} \quad (2.40)$$

A **sensibilidade** é a percentagem de classificações corretas no grupo “1-sucesso” da variável dependente, ou seja, a probabilidade de estimar o sucesso dado que o valor real observado é realmente sucesso, (ou seja, a capacidade do modelo indicar que o resultado do teste da criança exposta ao HIV é positivo dado que realmente o resultado do teste da criança exposta ao HIV é positivo).

$$SES = \frac{VP}{VP+FN} \quad (2.41)$$

A **especificidade** é a percentagem de classificações corretas no grupo “0-insucesso” ou seja, a probabilidade de estimar o insucesso dado que o valor real observado é realmente insucesso, (ou seja, a capacidade do modelo indicar que o resultado do teste da criança exposta ao HIV é negativo dado que realmente o resultado do teste da criança exposta ao HIV é negativo).

$$ESP = \frac{VN}{VN+FP} \quad (2.42)$$

Valor Preditivo Positivo - caracteriza à percentagem de verdadeiros positivos com relação ao total de predições positivas do modelo.

$$VPP = \frac{VP}{VP+FP} \quad (2.43)$$

Valor Preditivo Negativo - caracteriza à percentagem de verdadeiros negativos com relação ao total de predições negativas do modelo.

$$VPN = \frac{VN}{VN+FN} \quad (2.44)$$

Considera-se que um modelo com boas capacidades preditivas apresenta sensibilidade e especificidade superiores a 80%. Para percentagens entre 50% e 80% as capacidade preditivas são consideradas razoáveis. Abaixo de 50% as capacidades preditivas são consideradas medíocres (Maroco, 2007).

2.2.10.2. Curva ROC (*Receiver Operating Characteristic*)

De acordo com Hosmer e Lemeshow (1989), a análise ROC consiste em determinar-se uma curva que corresponde à razão entre a sensibilidade e um menos a especificidade (proporção de falsos positivos, isto é, proporção de incorrectamente classificados como tendo a característica a modelar, em relação ao total de sujeitos que não têm a característica a modelar). A curva ROC é desenvolvida num plano unitário de abcissa $1 - e$ e ordenadas, sobre o qual são traçados os diferentes pontos de corte das probabilidades relativas à sensibilidade e à especificidade. Deve-se escolher como ponto de corte aquele que mais se aproxima do canto superior esquerdo do plano unitário, ou seja, aquele que maximiza a sensibilidade e a especificidade. Quanto mais próximo do canto superior esquerdo o ponto de corte for, maior será o poder de discriminação do modelo.

Sobre o plano unitário é traçado também uma bissectriz $y = x$, que quanto mais próxima a curva estiver da bissectriz, menor é o poder de discriminação do modelo. A área debaixo da curva ROC, varia, obviamente, entre 0 e 1 e é utilizada para medir a eficiência de como o modelo consegue efetuar a classificação ou discriminação dos sujeitos. Se a área for igual a 0,5, o modelo não consegue discriminar os sujeitos com *versus* sem a característica. Quanto mais próximo a área estiver de 1

(isto é, da razão entre a sensibilidade igual a 0), maior é a capacidade do modelo para discriminar os sujeitos que apresentam a característica modelada (sucesso) dos indivíduos que não apresentam essa característica (insucesso).

Na tabela seguinte encontram-se os valores indicativos da área ROC que, servem como critério de aplicação geral para a descrição do poder discriminante apresentado pelo modelo de regressão logística:

Tabela 2.2.4: Valor indicativos da área ROC

Área ROC	Poder discriminante do modelo
0.5	Sem poder discriminante
]0.5;0.7[Discriminação fraca
]0.7;0.8[Discriminação aceitável
]0.8;0.9[Discriminação boa
≥ 0.9	Discriminação muito boa

Fonte: Hosmer e Lemeshow (1989)

2.2.11. Diagnóstico de *outliers* e de observações influentes

Segundo Mangas (2019), o diagnóstico da regressão logística, é baseado na **análise de resíduos** que permite identificar outliers e observações influentes na estimação do modelo ou na estimação de algum coeficiente em particular para tal as medidas usadas são similares às da regressão linear (resíduos, leverage, distância de Cook, e DfBetas). Os resíduos, não estandardizados, são definidos para a observação j como sendo a diferença entre y_j a probabilidade de sucesso estimada: $\hat{y}_j = e_j - \hat{y}_j$.

- **Resíduos de pearson**

Para obter resíduos estandardizados é necessário dividir os resíduos não-estandardizados pela estimativa do desvio padrão dos valores estimados:

$$e'_j = \frac{e_j}{\sqrt{\hat{\pi}_j(1-\hat{\pi}_j)}} \quad (2.45)$$

Estes resíduos apresentam média 0 e desvio-padrão aproximadamente 1. Para amostras de grande dimensão, a distribuição de e'_j é assintoticamente normal padrão. Assim, 95% dos valores $|e'_j|$ devem ser inferiores a 1,96 e qualquer observação com $|e'_j|$ superior a $1,96 \approx 2$ pode ser classificado como um outlier para o nível de significância de 5% (o quantil $Z_{0,95}=1,96$).

- **Leverage**

As influências das observações no modelo estimado serão estudadas através do Δx_j^2 em função das probabilidades estimadas, a observação com **Leverage** de aproximadamente 0 é pouco influente, enquanto que se for próximo do 1, então essa observação é importante mas as mais influentes são as que têm os valores de Δx_j^2 superiores a 3.84 para o modelo, porque a um nível de significância de $\alpha = 0.05$ o valor de $\chi_{0.95;(1)}^2 = 3.84$.

$$\Delta x_j^2 = \frac{(e'_j)^2}{1-b_j} \quad (2.46)$$

2.2.12. Verificação dos pressupostos do modelo logístico

- **Ausência da multicolinearidade**

Maia (2017), afirma que duas variáveis são ditas perfeitamente colineares quando há uma relação linear exata entre elas. Portanto a multicolinearidade refere-se apenas às relações entre as variáveis X , nesse caso, teríamos:

$$X_{ji} = \lambda_1 X_1 + \lambda_2 X_2 + \dots \dots \dots + \lambda_k X_k + \mu_i \quad (2.47)$$

Onde: $\lambda_1, \lambda_2, \dots \dots \lambda_k$ são constantes e μ_i é um termo de erro estocástico.

A existência de uma colinearidade exata entre duas ou mais variáveis independentes torna impossível a obtenção dos coeficientes dos parâmetros por MQO (mínimos quadrados ordinários). O problema é que a multicolinearidade torna muitas vezes as estimativas dos coeficientes dos parâmetros (β 's) insignificantes, já que cada um pressupõe, por definição, a variação em Y dada uma variação unitária em X , mantendo-se constantes as demais informações. Ou seja, se duas variáveis independentes são fortemente correlacionadas, será muito difícil haver variação em uma sem que haja em outra (Salvian, 2016).

- **Factor inflacionário da variância (VIF) e tolerância (TOL)**

Segundo Gujarati e Poster (2011), o VIF representa o quanto a variância de $\hat{\beta}_j$ está sendo inflacionada pela relação de multicolinearidade entre X_j e as demais variáveis independentes. Quando não houver relação entre as variáveis independentes ($R_j^2 = 0$) o VIF_j será igual à 1 e, à medida que nos aproximamos de uma relação exata ($R_j^2 = 1$), o VIF_j tenderá a infinito. Contudo o VIF é definido por:

$$VIF = \frac{1}{(1-R_j^2)} \quad (2.48)$$

A variância para cada estimador β_j , num modelo com k variáveis será dada por:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum X_j^2} \left(\frac{1}{1-R_j^2} \right) \quad (2.59)$$

Onde: $\hat{\beta}_j$ é coeficiente parcial (estimado) de regressão do regressor X_j , $R_j^2 = R^2$ na regressão de X_j sobre as $(k - 2)$ regressores remanescentes e $\sum X_j^2 = \sum (X_j - \bar{X}_j)^2$

Podemos escrever a equação (2.51) como:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum X_j^2} FIV \quad (2.50)$$

Segundo Gujarati e Poster (2011), modelo clássico, informa que, quanto maior a variabilidade de um regressor, menor a variância de seu coeficiente, supondo que os outros dois elementos sejam constantes e, portanto, maior a precisão com a qual esse coeficiente pode ser estimado.

Contudo podemos notar que o inverso de FIV é chamado de **tolerância** (TOL), ou seja:

$$TOL_j = \frac{1}{VIF_j} = (1 - R_j^2) \quad (2.51)$$

2.2.13. Seleção do melhor modelo

Segundo o Silveira (2021), os procedimentos automáticos de busca para a redução de variáveis na criação do modelo de regressão logística são nomeadamente: Forward, Backward, Stepwise. A seleção do modelo pode ser efectuada com o maior R^2 , mas também pode ser feita através dos critérios Akaike's Information Criterion (AIC) e Bayesian Information Criterion (BIC), em p variáveis explicativas.

- **A Forward:** consiste em inicialmente num modelo nulo, o procedimento adiciona, em cada etapa, a variável que daria o menor valor p da variável ainda não incluído no modelo. O procedimento para quando todas as variáveis foram adicionadas, ou quando nenhuma variável atende ao limite pré-especificado para o valor p.
- **Backward:** Considera-se inicialmente o modelo com todas as variáveis e por etapas elimina ou não cada uma delas, até que todas as variáveis restantes atendam a um limite especificado para seus valores p. Em cada etapa, a variável com o maior valor p é excluída.
- **O método Stepwise** é uma combinação dos dois métodos forward e backward.

De acordo com Santos (2016), na validação do desempenho dos modelos, o critério de informação de Akaike's Information Criterion (AIC) tem sido bastante utilizado para diferentes estruturas de covariâncias. O princípio de critério de informação de Akaike é selecionar uma combinação de variáveis exploratórias a modelos para a função de correlação que minimize o valor do AIC, ou seja quanto menor o valor de AIC melhor será o ajuste do modelo.

De acordo com o Velho (2015), este critério irá penalizar a log-verossimilhança de cada modelo, de forma a refletir o número do parâmetro que estão a ser estimados, permitindo a comparação dos modelos. A estimativa de AIC para um determinado modelo é dado por:

$$AIC = -2L(\theta) + 2k \quad (2.52)$$

Segundo o Oliveira (2015) O critério de informação Bayesiano (BIC), também chamado de critério de Schwarz é um critério de avaliação de modelos definido em termos da probabilidade a posteriori, sendo assim chamado porque forneceu um argumento Bayesiano para prová-lo. Segundo esses critérios, o melhor modelo será aquele que apresentar menor valor para AIC ou BIC.

As estatísticas BIC podem ser calculadas por:

$$BIC = -2L(\theta) + k \log(n) \quad (2.53)$$

Onde: $L(\theta)$ é o logaritmo de função verossimilhança do modelo e k é número de parâmetros do modelo e n é o número de observações em estudo.

2.2.14. Estudos Relacionados

Vários estudos tem sido feito, com diversas abordagens no que diz respeito aos factores de risco associados a seropositividade em crianças expostas ao HIV. Entretanto, alguns estudos são apresentados a seguir:

- Segundo o estudo feito pelo Nishimoto et al. (2005), mostraram a ocorrência do risco da transmissão materna infantil (TMI), estabelecida em 144 crianças, 14 delas foram classificadas como infectadas e o risco de transmissão foi estimada em 0.097, tendo sido menor nas mulheres com menos de 30 anos de idade. Na análise univariada o risco de TMI do HIV esteve associado de modo estatisticamente significativo com a idade materna, realização da consulta pré-natal, idade gestacional e amamentação materna.
- Siquira et al. (2020), no estudo efectuado sobre Análise hierarquizada dos determinantes da transmissão vertical do HIV: um estudo de caso-controle, mostraram a observação de 46,5% de mães com baixa escolaridade, 69,6% sem remuneração proveniente do trabalho e 35,7% recebendo bolsa familiar, mulheres com diagnóstico pós-parto e menos de 6 consultas de pré-natal apresentaram maior chance de transmissão vertical. As medidas profiláticas estiveram estatisticamente associadas à prevenção da transmissão ($p < 0,1\%$).
- Segundo o estudo feito pelo Figueredo (2023), com o objectivo de estimar a taxa de transmissão vertical (TV) em crianças expostas ao HIV no Hospital Universitário do Brasil, o mesmo foi composto por 725 crianças expostas das quais 672 crianças foram expostas e não infectadas e 53 infectadas, onde à maioria das gestantes tinham idade ≥ 20 anos em (86.9%) e (46.9%) tinham trabalho autônomo ou remunerado. Em relação assistência (86.3%) fizeram a CPN, (74.6%) fizeram a profláxia com terapia antirretroviral (Tarv) durante a gestação e (78.1%) tiveram parto cesário. Quanto as variáveis que tiveram associação com a variável de estudo foram: a escolaridade com (P-value=0.048), Profláxia do RN (P-value=0.002), CPN com (P-value=0.017) e conclui-se que a taxa de TV estimada neste estudo foi de 7,3% , evidenciando que ainda há fragilidade em adotar as intervenções recomendadas nos protocolos do MISAU na sua integridade para alcançar a redução e eliminação da TV do HIV.

- No estudo efectuado por Cruz (2012), composto por 1200 crianças expostas ao HIV, que contou com aplicação da análise logística uni e multivariada das características maternas, pós-parto e medidas preventivas para a identificação dos factores de risco para a taxa de transmissão materna infantil do HIV. Segundo o estudo a taxa de transmissão foi de 9,16% (IC 95% 7,4 – 10,9) e os factores de risco que permaneceram independentemente associados a transmissão do HIV são: não utilização dos antirretroviral durante a gestação (OR 7,84; IC 95% 4,1 – 15), parto normal (OR 2,02; IC 95% 1,2 – 3,4), prematuridade (OR 2,49; IC 95% 1,3 – 4,7) e aleitamento materno (OR 2,58; IC 95% 1,4 – 4,6).
- Carvalho (2015), realizou um estudo de coorte com 217 crianças expostas ao HIV, e tinha como objectivo do estudo analisar a vulnerabilidade das crianças expostas à transmissão vertical no período de 2000 a 2011 e obteve os seguintes resultados: as 117 mães de crianças expostas ao HIV tinham média de idade de 27 anos. Das 217 crianças (23,5%) adquiriram o vírus, (37,8%) não foram infectados e (38,7%) tiveram perda de seguimento. A análise de regressão logística as probabilidades mostraram um risco (2,38) vezes superior da criança ser infectada pelo HIV quando a mãe não faz o pré-natal, a não utilização dos medicamentos antirretrovirais durante a gestação aumentou em (9,65) vezes o risco de infecção da criança e o risco foi elevado em (7,55) vezes nas que receberam o aleitamento materno. A perda de seguimento foi identificado como um grave problema e estava associado a vulnerabilidade programática.
- O estudo efectuado por Silva (2004), que teve como objetivo avaliar a evolução de crianças nascidas de mães positivas para o HIV ou com AIDS no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, durante o período compreendido entre 1986 e 2001. Foram levantadas informações relativas a profilaxia pré-natal e da criança ao nascer, assim como à reversão sorológica, soropositividade e sobrevida. Dos 680 participantes, 67 (9,8%) se infectaram, 520 (76,5%) não se infectaram e 93 (13,7%) ficaram sem informação devido ao abandono de seguimento. Profilaxia durante a gestação ocorreu em 144 mulheres com o uso de uma droga (21,2%) e em 77 com a utilização de duas ou mais drogas (11,3%), não tendo se verificado em 459 gestantes (67,5%). Entre os recém nascidos, 205 (30,1%) receberam apenas AZT, 134 (19,7%) foram medicados com AZT+SMX/TMP e 341 (50,1%) não foram tratados.

CAPÍTULO 3

Material e Métodos

3.1 Material

No presente estudo, usou-se uma base de dados secundária, que resultou na administração de um questionário no ano de 2022 no período compreendido de 6 meses. Para o processamento desses dados, foi usado o pacote estatístico R Studio na versão 4.3.3, Excel e Word.

3.1.1 Local do estudo

O local escolhido para desenvolver o presente estudo foi o Distrito da Matola, isto é, concretamente em todas unidades sanitárias que respondem o mesmo Distrito. Entretanto, o Distrito da Matola localiza-se na capital da Província de Maputo, e tem limite a noroeste e a norte com o Distrito de Moamba, a oeste e sudeste com o Distrito de Boane, a sul e a leste com a Cidade de Maputo e a noroeste com o Distrito de Marracuene. O município tem uma área de 373 km^2 e a sua população de acordo com o censo de 2007 é de 672508 habitantes, estando dividido em três postos administrativos nomeadamente: Posto Administrativo de Infulene composto por 15 bairros, Posto Administrativo de Machava e Matola Sede com 13 e 13 bairros respectivamente.

Tabela 3.1: Referente as unidades sanitárias onde foi efectuado a recolha dos dados.

CS Ndlavela	CS Machava II	CS Malhampsene
CS Khongolote	CS Matola I	CS Bedene
CS Matola Gare	CS Matola II	CS Matola Santos
CS Boquisso	CS Tsalala	CS Liberdade
CS Khobe	CS Matola C	CS São Damanso
CS Unidade A	CS Machava I	CS Ligamo
CS Muhalaze		

3.1.2. Classificação do estudo

O método científico é fundamental para avaliar as pesquisas e seus resultados para que sejam aceites. Contudo, a pesquisa para seja científica, requer um procedimento formal, usando para isso método próprio e técnicas específicas.

A pesquisa é um processo permanente inacabado que é resultado de um inquérito ou exame minucioso, realizado com objectivo de resolver um problema, recorrendo a procedimentos científicos (Oliveira, 2011).

Segundo o Nascimento (2016), as pesquisas ou estudos podem ser classificados sub ponto de vista da sua natureza, aos métodos (ou abordagens metodológicas), quanto aos objectivos e procedimentos técnicos. Portanto, o presente estudo foi classificado da seguinte maneira:

- i. **Sub ponto de vista da sua natureza** – este estudo enquadra-se nas pesquisas aplicadas, porque com base nos resultados obtidos, os conhecimentos irão proporcionar soluções práticas para alguns problemas específicos que poderão estar assolando a população.
- ii. **Sub ponto de vista da abordagem** – este estudo enquadra-se nas pesquisas quantitativas porque recorre a uma linguagem matemática para descrever as causas de um fenómeno e traduz em números absolutos as opiniões ou informações da população participante no estudo, aplicando as técnicas estatísticas com objectivo de analisar as relações entre as variáveis e também buscando entender a influência de algumas variáveis explicativas sobre a variável dependente, aplicando a regressão logística.
- iii. **Sub ponto de vista dos seus objectivos** – trata-se de uma pesquisa explicativa porque foi realizado com vista a identificar os factores que determinam ou contribuem para a ocorrência da seropositividade em crianças expostas ao HIV.
- iv. **Sub ponto de vista dos procedimentos técnicos** – foi considerada uma pesquisa experimental, porque foi estabelecido um objecto de estudo (seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola) e seleccionam-se as variáveis que podem influenciá-la (sexo de criança exposta, tipo de parto, aleitamento da criança, PTV da criança, anos de tratamento tarv da mãe, a falta de alimentação da mãe, muitos parceiros sexuais, não adesão ao tratamento devido as crenças religiosas, sobre a toma dos medicamentos por parte da mãe, actualização da carga viral da mãe, abandono ao tratamento, inicio tardio do tarv da mãe durante a gestação, ocupação, uso da rede mosquiteira, fonte de água, tipo de casa e agregado familiar), definem-se os mecanismos e formas de controle e de observação que a variável produz no objecto.

3.1.3. População, amostra e amostragem

Para o presente estudo a população foi composta por todas as crianças expostas ao HIV de todas as unidades sanitárias do Distrito, com pelo menos 1º PCR² colhido durante o período de 1º semestre de 2022, assim sendo a população foi de N=1600 PCR's.

No que diz respeito a decisão do cálculo do tamanho da amostra, foi considerado um universo de N=1600 PCR's, com erro de 5%, nível de confiança de 95% (cujo o valor do $Z_{cr} = 1,96$), e tomando em consideração que a proporção populacional das crianças expostas por HIV devido aos factores de seropositividade é desconhecido, considerou-se 50% das crianças expostas ao HIV no Distrito da Matola.

$$n = \frac{Z^2 * p * q * N}{\varepsilon^2 * (N - 1) + Z^2 * p * q} = \frac{(1,96)^2 * 0,5 * 0,5 * 1600}{(0,05)^2(1600 - 1) + (1,96)^2 * 0,5 * 0,5} = 309$$

De acordo Mahaluça (2016), antes de obter uma amostra, é preciso definir os critérios que serão usados para seleccionar as unidades que vão compor essa amostra e de acordo com a técnica usada, tem-se um tipo de amostra.

A pesquisa contou com aplicação da Amostragem Probabilística Estratificada para a selecção dos processos individuais das crianças expostas ao HIV das unidades sanitárias onde efetuou-se a recolha dos dados. Tendo em conta que os indivíduos estão distribuídos em várias unidades sanitárias o tamanho da amostra foi obtida através do estudo proporcional que segue com a seguinte fórmula:

$$n_i = \frac{N_i}{N} * n$$

onde:

$i \rightarrow 1 \dots$ até 19 (que representa o total de 19 unidades sanitárias do Distrito);

$n \rightarrow$ Tamanho da amostra do estudo;

$n_i \rightarrow$ Tamanho da amostra do estrato por US em relação aos processos à avaliar;

$N_i \rightarrow$ Número de 1º PCR colhido por cada US durante o período em avaliação;

$N \rightarrow$ Número total de 1º PCR colhido de todas US's durante o período em avaliação.

² O PCR ADN para HIV é o teste viral (qualitativo) mais apropriado para as crianças menores de 18 meses e é o mais usado no país ou seja a realização precoce deste teste permite a administração também precoce de antiretrovirais, o que irá diminuir a morbimortalidade por esta doença, e melhorar qualidade de vida das crianças.

Tabela 3.2: Aplicação da amostragem probabilística estratificada para a determinação de amostra por cada unidade sanitária.

Centros de Saúde	N_i(1° PCR colhido por us)	n_i(Amostra por estrato)
CS Malhampsene	64	12
CS Bedene	39	8
CS Boquisso	88	17
CS Liberdade	54	10
CS Ligamo	21	4
CS Machava I	64	12
CS Matola C	71	14
CS Machava II	185	36
CS Matola Gare	170	33
CS Matola I	45	9
CS Matola II	209	40
CS Muhalaze	104	20
CS Matola Santos	23	4
CS Ndlavela	156	30
CS Nkobe	66	13
CS Damanso	69	14
CS Tsalala	59	11
CS Unidade A	23	4
CS Khongolote	90	18

3.1.4. Critérios de Inclusão e Exclusão

3.1.4.1. Inclusão

Crianças com exposição ao HIV/SIDA com resultado definitivo ou tenha feito pelo menos 1° PCR durante o 1° semestre de 2022.

3.1.4.2. Exclusão

Crianças com exposição ao HIV/SIDA que não estão compreendidos no 1° semestre de 2022 e também crianças sob outros cuidados na CCR.

3.1.5. Variáveis de estudo

Para o presente estudo foram consideradas as variáveis seguintes:

Tabela 3.3: Descrição das variáveis do estudo

Factores	Código da variável	Variáveis	Descrição da variável	Valor
Geral	r_test	Resultado do teste	Resultado do teste relativamente as crianças expostas	<i>Positivo</i> <i>Negativo*</i>
Demográfico	s_xo	Sexo	Representa o sexo das crianças	<i>Masculino*</i> <i>Feminino</i>
Clínicos	alei_mto	Aleitamento	O tipo de aleitamento usado pela criança exposta	<i>Materno</i> <i>Misto</i> <i>Artificial*</i>
	ptv_Cria	Ptv da criança	O tipo de medicamentos usados pela criança para prevenção de transmissão vertical	<i>NVP + AZT*</i> <i>AZT</i> <i>NVP</i> <i>Nenhum</i>
	ti_parto	Parto	O tipo de parto ocorrido na criança	<i>Normal</i> <i>Cesariano*</i>
	a_trat_mae	Tratamento Tarv da mãe	Corresponde aos anos de tratamento Tarv que as mães seropositivas tem com a doença	<i>< 3 anos*</i> <i>3 - 6 anos</i> <i>> 6 anos</i>
	c_viral	Carga viral	Actualização da última carga viral da mãe	<i>Suprimida*</i> <i>Não Suprimida</i> <i>Nenhuma</i>
	a_bandno	Abandono	Registo de abandono da mãe ao tratamento Tarv	<i>Sim</i> <i>Não*</i>
	i_ncio	Início Tarv tardio	Representa o início Tarv tardiamente da mãe durante a gestação	<i>Sim</i> <i>Não*</i>
	a_metcao	Alimentação	Corresponde a falta de alimentação por parte das mães seropositivas	<i>Sim</i> <i>Não*</i> <i>As vezes</i>
Sociais	r_giao	Religião	Refere a situação de não adesão ao tratamento por conta das crenças religiosas	<i>Sim</i> <i>Não*</i>
	m_parcer	Parceiros	É a relação da existência de outros parceiros sexuais da mãe seropositiva	<i>Sim</i> <i>Não*</i>
	a_famly	Agregado familiar	Corresponde ao agregado familiar	<i><=3 AF*</i> <i>4 - 6 AF</i> <i>>=7 AF</i>
	c_sa	Casa	Refere ao tipo de casa usada	<i>Bloco</i> <i>Canço*</i>

	f_aga	Fonte de água	Corresponde ao tipo de distribuição de água usada	Água Canalizada Poço* Água + de 50 m
	f_quiter	Mosquiteira	Se usa alguma rede mosquiteira	Sim * Não
	m_dimnto	Medicamentos	Se a mãe tem tomado regularmente os medicamentos antirretrovirais	Sim toma* Não toma Falha na toma
Económica	o_pcao	Ocupação	Representa a fonte de rendimento	Emprego* Auto - Emprego Ag. Subsistência Nenhum

* Representam as categorias de referência para cada variável do estudo.

3.2. Métodos

3.2.1. Técnicas da estatística usadas (procedimentos do estudo)

- **Teste de independência**

No presente trabalho foi usado o teste de *Qui-quadrado de pearson* (χ^2) para avaliar associação entre a variável dependente das independentes.

- **Para testes de significância dos coeficientes**

Foram usados os testes de razão de verossimilhança e wald para analisar a significância do modelo e os efeitos dos coeficientes. Entretanto o teste de razão de verossimilhança foi usada para testar os efeitos gerais do modelo, isto é, a verificação de pelo menos se um dos coeficientes é estatisticamente significativo e diferente de zero e wald para testar se cada um dos coeficientes é feita de forma individual e estatisticamente significativo ao nível de significância de 5%.

- **Odds (OR) razão de chance**

Depois do teste de significância dos coeficientes estimados do modelo, fez-se o cálculo de *Odds (OR)* e respectivo IC à 95%.

- **Pressupostos do modelo**

Quanto aos pressupostos do modelo foi avaliado, ausência de outliers dos resíduos assim como os seus pontos influentes, ausência de multicolinearidade entre as variáveis através do teste do VIF³ que deve ser abaixo de 10.

- **Quanto avaliação da qualidade do ajuste do modelo estimado**

Avaliação da qualidade do ajuste do modelo foi feito usando *Pseudo R²* de “Nagelkerke”, onde indica a percentagem da variância do modelo, ou seja, em que nível a variável dependente esta sendo explicada pelas variáveis explicativas.

- **Classificação, sensibilidade, especificidade e curva de ROC**

No presente estudo foi calculado o valor da classificação, sensibilidade e especificidade do modelo através da matriz cruzada e foi considerado o coorte de classificação de 0,5 (50%). Também foi avaliada a curva de *ROC* que tem como objectivo buscar a relação entre a taxa dos verdadeiros positivos do modelo e a taxa dos falsos positivos preditos.

Por último foram calculadas as probabilidades de cada criança de poder ter o resultado positivo ou negativo em função das variáveis significativas do modelo escolhido, no intervalo de confiança de 95%. Em relação ao calculo das probabilidades foi usada uma amostra de n=2 e contou com aplicação da amostragem probabilística aleatória simples para gerar os números que haveriam de compor a amostra.

³ Variance Inflation Factor

CAPÍTULO 4

Resultados e Discussão

Neste capítulo, procura-se apresentar os resultados do estudo tendo em conta os dados colectados através do instrumento de recolha de dados adoptado no mesmo e fazer uma discussão com resultados encontrados em estudos similares, através de uma abordagem teórica aquelas que são as ideias ou posicionamento dos diversos autores nesta matéria em análise.

4.1. Apresentação dos resultados

4.1.1 Descrição do perfil da amostra

A Figura 4.1, representa a distribuição das crianças expostas ao HIV/SIDA por sexo, onde das 300 crianças estudadas, 45% são do sexo masculino e 55% são do feminino.

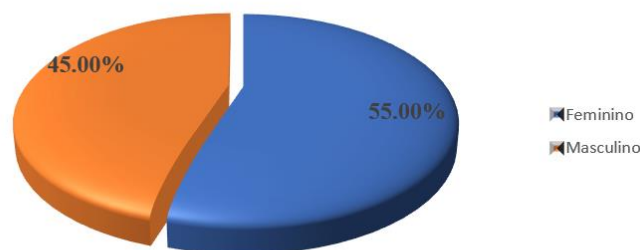


Figura 4.1: Distribuição das crianças expostas por sexo

4.1.2 Perfil sociodemográfico e económico das crianças expostas ao HIV/SIDA no Distrito da Matola

Neste ponto, os dados estão apresentados por sessões referente as características sociodemográficas e económicas das crianças expostas ao HIV/SIDA no Distrito da Matola e estão divididas em: aspectos demográficos, clínicos, sociais e económicos.

Perfil sociodemográfico e económico das crianças expostas ao HIV/SIDA no Distrito da Matola

Em relação ao a variável sexo, observou-se que 89% das crianças expostas ao HIV eram negativas e 11% delas eram positivas, tendo em conta que os maculinos tinham uma percentagem de 45% e 55% para o sexo feminino.

Quanto avaliação da variável ocupação da mãe seropositiva, 89% eram negativas contra 11% positivos. Olhando para as modalidades da variável, apresentaram maior prcentage de positividade, as mães que não tinham nenhuma ocupação em 6%, seguindo o emprego e auto-emprego com percentagens de 2% respectivamente.

Na análise bivariada do teste chi-quadrado de pearson (χ^2), verificou-se que nem o sexo e a ocupação da mãe seropositiva podem ser associados aos factores da seropositividade nas crianças expostas ao HIV/SIDA, visto que $p - valor$ para as duas variáveis é superior a 5%.

Tabela 4.1: Perfil sociodemográfico e económico das crianças expostas ao HIV/SIDA no Distrito da Matola

Variáveis	Categorias	Crianças expostas		Teste de χ^2 (P - Valor)
		Negativo	Positivo	
Sexo	Feminino	145 (48%)	20 (7%)	0.7696
	Masculino*	121 (41%)	14 (4%)	
Ocupação da mãe	Agricultura de Subsistência	14 (5%)	4 (1%)	0.4856
	Auto - Emprego	68 (23%)	7 (2%)	
	Emprego*	58 (19%)	7 (2%)	
	Nenhuma	126 (42%)	16 (6%)	

Perfil social das crianças expostas ao HIV/SIDA no Distrito da Matola

Olhando para avaliação social das crianças expostas observou-se que 2% das mães seropositivas tem dependência das creanças religiosas para adesão do tratamento contra 98 % sem dependência, contudo a maior percentagem de positividade foi registada nas mães sem dependência em 9%.

Quanto a existência de outros parceiros sexuais da mãe seropositiva, os dados colhidos revelaram que 88% delas afirmaram ter relações sexuais com mesmo parceiro (marido), tendo uma percentagem de positividade de 9% contra 12% que afirmaram ter outros parceiros sexuais, correspondendo assim à 2% de positividade.

Em relação ao agregado familiar, observou-se que maior parte das crianças expostas viviam em famílias composta por elementos de entre 4 – 6 AF, com uma percentagem de 48% e positividade de 6%. Também é possível observar uma maior percentagem das crianças em famílias com elementos ≥ 7 AF, com 38% e positividade de 4%.

No que tange o tipo de casa usada, observou-se que a maior parte viviam em casas modernas feitas de bloco com um percentagem de 98%, com uma taxa de positivos de 11%. Olhando para o tipo de fonte de água, 97% tinham água canalizada e apenas 3 % ainda usavam a água de fonte do poço.

Observou-se também que maior percentagem das mães usavam as redes mosquiteiras nas suas casas com 91% contra 9% que não usava. No que diz respeito a toma regular dos medicamentos antirretrovirais por parte das mães, constatou-se 89% afirmou tomar de forma regular, 3% não tomava e 8% afirmou ter as vezes falha na toma.

Na análise bivariada do teste chi-quadrado de pearson (χ^2), verificou-se que no perfil social a religião (associada a dependência das crenças religiosas para não adesão do tratamento) e os medicamentos podem ser associados aos factores da seropositividade nas crianças expostas ao HIV/SIDA, visto que p – valor é inferior a 5%.

Tabela 4.2: Perfil social das crianças expostas ao HIV/SIDA no Distrito da Matola

Variáveis	Categorias	Crianças expostas		Teste de χ^2 (P - Valor)
		Negativo	Positivo	
Religião	Sim	0 (0.00%)	7 (2%)	< 0.0001*
	Não*	266 (89%)	27 (9%)	
Parceiros	Sim	28 (10%)	7 (2%)	0.1506
	Não*	238 (79%)	27 (9%)	
Agregado familiar	≤ 3 AF*	36 (12%)	5 (2%)	0.9826
	4 - 6 AF	127 (42%)	16 (6%)	
	≥ 7 AF	103 (34%)	13 (4%)	
Casa	Bloco	263 (87%)	32 (11%)	0.1842
	Canço*	3 (1%)	2 (1%)	
Fonte de água	Água Canalizada	260 (87%)	31 (10%)	0.07503
	Poço*	5 (2%)	2 (1%)	
	Água + de 50 m	1 (0%)	1 (0%)	
Mosquiteira	Sim *	243 (81%)	32 (10%)	0.8261
	Não	23 (8%)	2 (1%)	
Medicamentos	Sim toma*	245 (82%)	22 (7%)	< 0.0001*
	Não toma	7 (2%)	3 (1%)	
	Falha na toma	14 (5%)	9 (3%)	

Perfil clínico da mãe e das crianças expostas ao HIV/SIDA no Distrito da Matola

No perfil clínico da mãe e das crianças expostas ao HIV/SIDA, observou-se que a maior percentagem das crianças expostas usavam aleitamento materno em 95%, seguido por aleitamento misto com 5%, tendo tido uma positividade de 10% e 1% respectivamente. No que diz respeito a prevenção de transmissão vertical a medicação de NVP + AZT apresentou a maior percentagem com 86% e positividade de 9%.

Observou-se também que 96 % dos partos foram normais e com positividade de 34% contra 4% de parto cesariano e com 0% de positividade.

Quanto a variável que avalia os anos de tratamento tarv da mãe, observou-se que a maior percentagem estava compreendida em faixa de < 3 anos com 54% e positividade de 5%. Também foi possível observar que as mães que tem > 6 anos de tratamento tem menor positividade.

Em relação a carga viral, constatou-se que 26% das mães na ultima actualização não tinham carga suprimida e 8% não tinham nenhuma informação reactivamente a sua carga viral actualizada. No que tange a variável abandono ao tratamento 99% das mães afirmaram não ter abandonado o tratamento desde que conheceram o seu seroestado.

De acordo com os dados relativos ao início tardío do tarv durante a gestação 98% mostraram ter iniciado no tempo recomendado o tratamento, tendo tido uma positividade de 10%. Quanto avaliação das dificuldades tidas para alimentação por parte das mães, 86% afirmou não ter dificuldades, 1% sem dificuldades e 3% afirmou que as vezes tem tido dificuldades

Na análise bivariada do teste chi-quadrado de pearson (χ^2), verificou-se que o ptv da criança, carga viral da mãe, abandono ao tratamento, início tardío do tarv durante a gestação e a variável alimentação podem ser associados aos factores da seropositividade nas crianças expostas ao HIV/SIDA, visto que $p - valor$ é inferior a 5%.

Tabela 4.3: Perfil clínico da mãe e das crianças expostas ao HIV/SIDA no Distrito da Matola

Variáveis	Categorias	Crianças expostas		Teste de χ^2 (P - Valor)
		Negativo	Positivo	
Aleitamento	Materno	254 (85%)	31 (10%)	0.4478
	Misto	11(4%)	3 (1%)	
	Artificial*	1 (0%)	0 (0%)	
Ptv da criança	NVP + AZT*	232 (77%)	27 (9%)	0.00695*
	AZT	19 (6%)	2 (1%)	
	NVP	13 (4%)	2 (1%)	

	Nenhum	2 (1%)	3 (1%)	
Parto	Normal	255(85%)	34 (11%)	0.4693
	Cesariano*	11 (4%)	0 (0.00%)	
Tratamento Tarv da mãe	< 3 anos	147 (49%)	14 (5%)	0.2291
	3 - 6 anos	86 (29%)	13 (4%)	
	> 6 anos	33 (11%)	7 (2%)	
Carga viral	Suprimida*	186 (62%)	11 (4%)	0.00038*
	Não Suprimida	63 (21%)	16 (5%)	
	Nenhuma	17 (6%)	7 (2%)	
Abandono	Sim	1 (0%)	4 (1%)	0.0003*
	Não*	265 (89%)	30 (10%)	
Início Tarv tardio	Sim	3 (1%)	3 (1%)	0.0179*
	Não*	263 (88%)	31 (10%)	
Alimentação	Sim	0 (0.00%)	4 (1%)	< 0.0001*
	Não*	260 (87%)	27 (9%)	
	As vezes	6 (2%)	3(1%)	

4.1.3 Análise da associação das variáveis do estudo

Na análise bivariada com o teste de *qui-quadrado de pearson* das variáveis do estudo, constatou-se que 7 variáveis estão associadas a seropositividade em crianças expostas ao HIV/SIDA no distrito da Matola, nomeadamente:

1. Religião (Não adesão ao tratamento por conta de crenças religiosas);
2. Medicamentos (toma regular dos medicamentos);
3. Ptv da criança (prevenção de transmissão vertical da criança);
4. Carga viral (actualização da última carga viral da mãe);
5. Abandono (abandono ao tratamento tarv);
6. Início tarv tardio (início tardio do tarv durante a gestação);
7. Alimentação (se há falta de alimentação por parte das mães seropositivas).

Na tabela 4.2 e 4.3 são apresentados variáveis independentes significativos, porém para análise de Odds Ratio e seus intervalos de confiança a 95% foram considerados as variáveis significativos da tabela 4.4, onde observou-se que uma mãe que tem dificuldades em se alimentar por vezes tem 9 vezes a chance de ter uma criança com resultado do teste positivo em relação a que se alimenta com o intervalo de confiança de 95% ($OR= 8.95$; IC 95% 1.74; 45.86).

Em relação a toma regular dos medicamentos, uma mãe que tem tido falha na toma dos medicamentos tarv tem 9 vezes de chance de ter uma criança com resultado do teste positivo comparando com uma mãe que toma ($OR= 9.042$; IC 95% 2.49; 32.82).

Quanto actualização da carga viral, uma mãe com carga viral não suprimida tem 4 vezes a chance de ter uma criança com resultado do teste positivo em relação com uma mãe com carga viral suprimida com o intervalo de confiança de 95% ($OR=3.74$; IC 95% 1.23; 11.31). Observa-se também que mães sem nenhum registo de carga viral tem 7 vezes a chance de ter uma criança com resultado do teste positivo comparando com uma mãe com carga viral suprimida, tendo um intervalo de confiança de 95% ($OR=7.09$; IC 1.70; 29.56).

No abandono ao tratamento tarv, observa-se que uma mãe que tenha abandonado o tratamento tem 44 vezes a chance de ter uma criança com resultado do teste positivo comparando com uma mãe que nunca abandonou o tratamento, com um intervalo de confiança de 95% ($OR=44.35$; IC 3.20; 613.85).

Tabela 4.4: Intervalo de confiança a 95% de Odds Ratio para categorias de variáveis significativas

Variável	Categoria	OR (IC 95%)	P-valor
Ptv da criança	NVP + AZT*	Referência	-
	AZT	1.47 (0.24; 8.91)	0.674954
	NVP	2.51 (0.47; 13.22)	0.277472
	Nenhum	12.16 (0.74; 197.6)	0.079083
Alimentação	Sim	Inf (0.000;inf)	-
	Não*	Referência	-
	As vezes	8.95 (1.74; 45.86)	0.008537*
Religião	Sim	Inf (0.000;inf)	-
	Não*	Referência	-
Medicamentos	Sim toma*	Referência	-
	Não toma	0.41 (0.0019; 89.77)	0.751210
	Falha na toma	9.042 (2.49; 32.82)	0.000815*
Carga viral	Suprimida*	Referência	-
	Não Suprimida	3.74 (1.23; 11.31)	0.019241*
	Nenhuma	7.09 (1.70; 29.56)	0.007098*
Abandono	Sim	44.35 (3.20; 613.85)	0.004673*
	Não*	Referência	-
Início Tarv tardio	Sim	8.99 (0.85; 94.29)	0.066918
	Não*	Referência	-

4.1.4 Análise de regressão logística e estimativa de modelo de regressão logística para os factores associados a seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola

Para o presente estudo foram construídos dois modelos de regressão logística, o primeiro modelo contendo todas as variáveis do estudo (18 variáveis) e o segundo modelo feito com base nas variáveis

significativas da análise bivariada com recurso ao teste de qui-quadrado de pearson (χ^2). Entretanto esses dois modelos depois de serem criados foram comparados onde o melhor modelo foi considerado no estudo.

Construção e comparação dos modelos de regressão logística (MRL)

De acordo com a Tabela 4.5 tendo em conta o método de comparação dos modelos utilizando o AIC e BIC, constatou-se que o melhor modelo é o 2 porque apresenta o menor valor do AIC e BIC em comparação ao modelo1.

Tabela 4.5: Comparação dos modelos construídos através do AIC e BIC

Modelo	Modelos Construídos	AIC	BIC
1	Modelo1	158.74	266.15
2	Modelo2	146.11	194.26

Análise de pressupostos do modelo2 de regressão logística, verificação valores influentes (outliers) e multicolinearidade

Na análise dos resíduos do modelo estimado constou-se não existirem pontos influentes na estimação do modelo, ou seja os dados não apresentaram nenhum outliers ao longo das observações. Dai que os resíduos variaram de -2.066 a 2.826 como valores mínimo e máximo respectivamente.

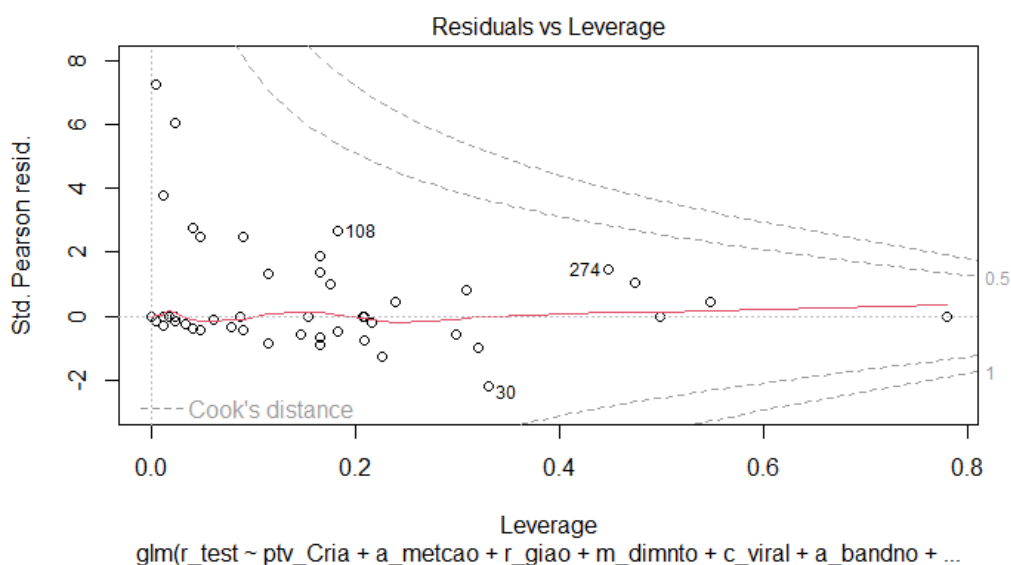


Figura 4.2: Análise de outliers do modelo2

Em relação a verificação de multicolinearidade dos dados, observou-se o VIF e a Tolerância. E constatou-se que para todas as variáveis inseridas no modelo apresentaram um VIF inferior a 10, concluindo que não existe multicolinearidade entre as variáveis.

Tabela 4.6: Análise de multicolinearidade do modelo2

Variáveis	Multicolinearidade	
	Tolerância	VIF
Ptv da criança	0.9779	1.0225
Alimentação	0.9895	1.0106
Religião	1.0000	1.0000
Medicamentos	0.9531	1.0492
Carga viral	0.9686	1.0324
Abandono	0.9632	1.0382
Início tardio	0.9573	1.0445

Teste de razão de verossimilhança

Este teste avalia a significância dos coeficientes estimados simultaneamente, verificando se o modelo é globalmente significativo. A razão de verossimilhança fundamenta-se através da função anova, que foi rejeitada a hipótese nula que dizia que o modelo não é estatisticamente significativo, visto que o valor do p- valor é inferior a 5% e logo conclui-se que pelo menos uma variável independente do modelo2 influência na seropositividade nas crianças exposta ao HIV/SIDA no distrito da Matola.

Tabela 4.7: Teste de razão de verossimilhança para avaliação global do modelo

Variáveis	Df	Pr (> chisq)
Ptv da Criança	3	0.3397
Alimentação	2	0.0007*
Religião	1	0.0004*
Medicamentos	2	0.0082*
Carga viral	2	0.0083*
Abandono	1	0.0032*
Início tardio	1	0.1084

O presente estudo apresentou o pseudo- R^2 "Nagelkerke" em 52.08% e mostrou que há dificuldades por vezes na alimentação, a falha na toma regular dos medicamentos, o facto de não ter nenhum registo da actualização da carga viral, carga não suprimida e ter registo de abandono do tratamento, essas são variáveis previsoires para determinar os factores associados a seropositividade das crianças expostas ao HIV.

Tabela 4.8: Modelo de regressão logística para a determinação dos factores associados a seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola

Variável	Coef.	Test-wald	P-valor	OR (IC 95%)
<i>Intercept</i>	-3.9570	0.4904	< 0.0001*	0,19 (0.007; 0.049)
ptv_CriaAZT	0.3857	0.9197	0.6749	1.47 (0.24; 8.91)
ptv_CriaNenhum	2.4982	1.4227	0.0790	12.16 (0.74; 197.6)
ptv_CriaNVP	0.9205	0.8476	0.2774	2.51 (0.47; 13.22)
a_metcaos vezes	2.1921	0.8335	0.0085*	8.95 (1.74; 45.86)
a_metcaosim	20.8350	2475.69	0.9932	Infinity
r_giaosim	20.6348	2245.75	0.9926	Infinity
m_dimntoFalha na toma	2.2019	0.6578	0.0008*	9.042 (2.49; 32.82)
m_dimntoNão toma	-0.8679	2.7374	0.7512	0.41 (0.0019; 89.77)
c_viralNao Suprimida	1.3204	0.5641	0.0192*	3.74 (1.23; 11.31)
c_viralNenhuma	1.9598	0.7280	0.0070*	7.09 (1.70; 29.56)
a_bandnosim	3.7922	1.3406	0.0046*	44.35 (3.20; 613.85)
i_nciosim	2.1966	1.1989	0.0669	8.99 (0.85; 94.29)

As Categorias de referência para cada variável são: Ptv da criança (refer: NVP + AZT*), Alimentação (refer: Não*), religião (refer:não*), medicamento (refer: sim toma*), carga viral (refer: suprimida*) e início tardio (refer: não*).

Teste de ajustamento do modelo de regressão logística através de teste de Hosmer e Lemeshow

Com objectivo de testar a qualidade do ajuste do modelo, utilizou-se o teste de Hosmer e Lemeshow para avaliar se o modelo obtido pode explicar adequadamente os dados observados e os previstos pela variável resultado do teste da criança exposta, sob as seguintes hipóteses:

H_0 : O modelo se ajusta aos dados

H_1 : O modelo não se ajusta aos dados

Mediante aos dados, observou-se que o p-valor (< 0.0001) associado ao teste é menor que 5% , o que significa que há evidências suficientes para rejeitar a hipótese nula, que afirma o modelo logístico ajusta-se aos dados. Porém segundo Hosmer & Lemeschow (2000), afirma que um modelo pode ser aceitável se tiver um AUC igual ou superior de 0.7, assim sendo o presente modelo pode ser usado

para a previsão da seropositividade das crianças expostas ao HIV/SIDA no Distrito da Matola, visto que tem uma estimativa de AUC de 0.87.

Tabela 4.9: Teste de Hosmer e Lemeshow

Chi-Squared	Grau de liberdade	P-Valor
300	3	< 0.0001

Curva ROC (*Receiver Operating Characteristic Curve*) do modelo de regressão logística

Em relação a Curva ROC para a determinação da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola, os resultados encontrados mostram ter capacidade preditiva boa ao apresentar um AUC de 0,877 (87.7%) segundo o quadro de classificação do Hosmer & Lemeshow (2000).

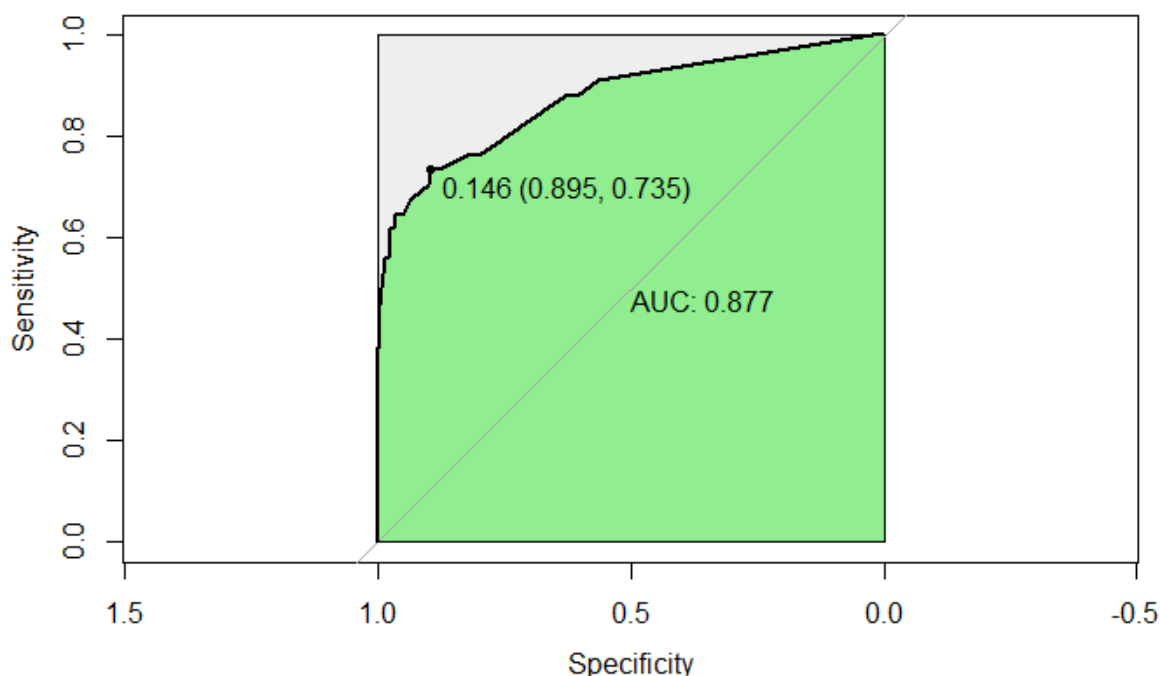


Figura 4.3: *Receiver Operating Characteristic Curve (ROC)* do modelo de regressão logística para a determinação da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola.

Classificação, sensibilidade e especificidade do Modelo de Regressão Logística

Classificação representa a proporção das predições correctas do modelo sobre o total e foi de 93%. A **sensibilidade** é a capacidade do modelo indicar que o resultado do teste da criança exposta ao HIV é positivo dado que realmente o resultado do teste da criança exposta ao HIV é positivo e foi de 89%. Quanto a **especificidade** é a capacidade do modelo indicar que o resultado do teste da criança exposta ao HIV é negativo dado que realmente o resultado do teste da criança exposta ao HIV é negativo e foi de 94%.

Tabela 4.10: Matriz de cruzada do modelo de regressão logística a determinação da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola.

		Valor Observado	
		Positivo	Negativo
Valor Estimado	Positivo	16	18
	Negativo	2	264

Expressão da regressão logística do estudo

Nesta secção são considerados as variáveis que no modelo2 escolhido permaneceram estatisticamente significante por apresentarem menor p-valor e por estarem associado com a variável dependente, nomeadamente: Ptv da criança, alimentação da mãe positiva, religião que esta associado as crenças relegiosas para não aderir ao tratamento, toma de medicamentos de forma regular, carga viral, abandono e início tardio do tarv.

Portanto, em seguida é apresentado o modelo logístico com a capacidade de prever a probabilidade de ocorrência da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola.

Tabela 4.11: Modelo logístico para para a determinação da seropositividade em crianças expostas ao HIV/SIDA no Distrito da Matola

Variável	Categoria	Coef.
Ptv da criança	<i>Intercept</i>	-3.9570
	Ptv_CriaAZT	0.3857
	Ptv_CriaNenhum	2.4982
	Ptv_CriaNVP	0.9205
Alimentação	a_metcaos vezes	2.1921
	a_metcaosim	20.8350
Religião	r_giaosim	20.6348
Medicamentos	m_dimntoFalha na toma	2.2019

	m_dimntoNão toma	-0.8679
Carga viral	c_viralNao Suprimida	1.3204
	c_viralNenhuma	1.9598
Abandono	a_bandnosim	3.7922
Início tardio do tarv	i_nciosim	2.1966

$$\hat{\pi} = \frac{e^{-3.9570 + \sum_{n=1}^7 \text{variável} \times \text{coeficiente}}}{1 + e^{-3.9570 + \sum_{n=1}^7 \text{variável} \times \text{coeficiente}}}$$

Onde:

$\hat{\pi}$: representa a probabilidade da criança exposta ao HIV ter o resultado positivo/negativo.

Aplicação da expressão da regressão logística do estudo

Para a aplicação da expressão da regressão logística do estudo, fez-se uma selecção de 2 crianças expostas ao HIV utilizando o método de selecção da amostragem aleatoria simples no qual foram escolhidas as seguintes posições: 38 e 147 do total das 300 observações estudadas.

$$\hat{\pi}(x_{38}) = \frac{e^{-3.957 + 1 \times 1,3204}}{1 + e^{-3.9570 + 1 \times 1,3204}} = 0,06$$

$$\hat{\pi}(x_{147}) = \frac{e^{-3.957 + 2 \times (-0,8679) + 1 \times 1,3204 + 1 \times 3,7922}}{1 + e^{-3.957 + 2 \times (-0,8679) + 1 \times 1,3204 + 1 \times 3,7922}} = 0,57$$

Interpretação:

Posição 38, trata-se de uma criança que a mãe não tem falta de alimentação, sem dependência das crenças para aderir o tratamento, que toma a medicação regularmente, não tem actualiza a sua carga viral e sem histórico de abandono ao tratamento tarv, a sua criança tem 6% de chance de testar positivo ao HIV.

Posição 147, trata-se de uma criança que a mãe tem falta de alimentação, sem dependência das crenças para aderir o tratamento, que não toma a medicação regularmente, não actualiza a sua carga viral, com histórico de abandono, a sua criança tem 57% de chance de testar positivo ao HIV.

4.2. Discussão dos resultados

Com o presente estudo foi possível observar que a maior parte dos casos de PCR's positivos não são reportados a nível das unidades sanitárias, fazendo com que as taxas de contaminação das crianças expostas ao HIV continuem a se alastrar na província, nos bairros e hospitais. Por outro lado constatou-se que há fragilidades na harmonização e fiscalização dos instrumentos de registo, recolha e seguimento da criança exposta ao HIV, visto que essa problemática está aliada a falta de comunicação entre os parceiros que apoiam as unidades sanitárias, que acabam fazendo um trabalho particular em busca dos objectivos da organização. Portanto há uma necessidade de se intensificar estudos do gênero com vista a trazer mais factores que possivelmente possam estar aliados a seropositividade das crianças expostas ao HIV.

Em relação aos resultados obtidos no estudo tem a se destacar o seguinte:

As dificuldades para uma mãe positiva se alimentar apresentou uma associação significativa com a variável dependente do estudo, aonde destacou-se que uma mãe que tem dificuldades em se alimentar tem 9 vezes a chance de ter uma criança com resultado do teste positivo em relação a que se alimenta, em consonância à estes achados, um estudo de (Kreitchmann, 2003) afirmou que as pessoas seropositivas devem seguir rigorosamente algumas regras de alimentação para ajudar o seu tratamento e melhorar a sua qualidade de vida porque tendem à apresentar diversos problemas secundários ocasionados pela doença.

Em relação a toma regular dos medicamentos também destacou-se como uma das variáveis significativas do estudo. Esta variável foi destacada também como factor de associação significativa de transmissão vertical no estudo de Cruz (2012), onde mostrou que a não utilização dos antirretroviral durante a gestação aumenta o risco de positividade em crianças expostas. Outros achados corrobora com estudos de (Santos e Souza; Silva et. 2016), afirmam que adesão ao tratamento assim com o uso de medicamentos se destaca entre os principais maiores desafios da atenção às pessoas vivendo com HIV/SIDA e devem ter para melhorar a sua qualidade de vida.

O pvt da criança também mostrou-se significativo segundo os dados analisados. Porém essa questão do pvt da criança é uma situação ligada directamente aos cuidados da mãe perante a criança, visto que trata-se de uma medicação para prevenção do HIV.

A variável carga viral esteve associado significamente a variável de estudo, onde destacou-se que uma mãe com carga viral não suprimida tem 4 vezes a chance de ter uma criança com resultado do

teste positivo em relação com uma mãe com carga viral suprimida . Observou-se também que mães sem nenhum registo de carga viral tem 7 vezes a chance de ter uma criança com resultado do teste positivo comparando com uma mãe com carga viral suprimida. Desta forma, estes achados foram observados no estudo de Bertoldo, que afirma quanto menor for a carga viral de mãe, menor será a possibilidade de transmissão vertical. (Bertoldo, 2014)

A perda de seguimento neste caso específico o abandono foi considerada uma variável significativa no presente estudo. Desta forma alguns achados foram encontrados no estudo de coorte de Carvalho (2015), onde o autor considera que a perda de seguimento foi identificado como um grave problema e estava associado a vulnerabilidade programática (factores clínicos e sociais).

O início tardio do tarv durante o período de gestação também foi um factor determinante da seropositividade das crianças. O carvalho (2015) no seu estudo afirma que a não utilização dos medicamentos antirretrovirais durante a gestação aumentou em (9,65) vezes o risco de infecção da criança ao vírus de HIV. Por outro lado Cunga et al. (2022), no seu estudo afirma que o diagnostico tardio e o desconhecimento das formas de transmissão do HIV/SIDA podem contribuir para o aumento do número de casos em neonatais, sobretudo pela carga viral materna elevada.

Capítulo 5

Conclusões e Recomendações

5.1. Conclusões

Há seropositividade em crianças expostas ao HIV/SIDA é um factor que ocorre devido a negligência de informações por parte das gestantes vivendo com HIV/SIDA, contribuindo assim a não adesão as práticas de prevenção consequentemente aumentando desta forma o risco de transmissão vertical na população pediátrica. Portanto levando em consideração os objectivos estabelecidos e de acordo com os resultados deste estudo, conclui-se que:

- Relativamente ao perfil sócio-demográfico das crianças expostas ao HIV/SIDA, 45% eram do sexo masculino e 55% eram do sexo feminino;
- Em análise da associação entre as variáveis sociodemográfico, clínicos, sociais e económicas foi feita análise bivariada usando o teste qui-quadrado com todas as variáveis independentes em função da dependente onde constatou-se que as variáveis que demonstraram ter alguma associação foram as seguintes: medicamentos, alimentação, início tarv tardio, abandono, carga viral e pTV da criança. Portanto as mesmas variáveis foram analisadas usando a regressão logística onde os factores que melhor se ajustam para a previsão foram: alimentação, religião, medicamentos, carga viral e abandono.
- Tendo as variáveis significativas do modelo foi possível estimar a equação que possa prever a probabilidade do resultado do teste para as crianças expostas ao HIV/SIDA no distrito da Matola.
- Relativamente ao *OR* em função das variáveis significativas destacou-se o seguinte:
 - Uma mãe que tem dificuldades em se alimentar por vezes tem 9 vezes a chance de ter uma criança com resultado do teste positivo em relação a que se alimenta com o intervalo de confiança de 95% ($OR = 8.95$; IC 95% 1.74; 45.86).
 - Em relação a toma regular dos medicamentos, uma mãe que tem tido falha na toma dos medicamentos tarv tem 9 vezes de chance de ter uma criança com resultado do teste positivo comparando com uma mãe que toma ($OR = 9.042$; IC 95% 2.49; 32.82).

- Quanto actualização da carga viral, uma mãe com carga viral não suprimida tem 4 vezes a chance de ter uma criança com resultado do teste positivo em relação com uma mãe com carga viral suprimida com o intervalo de confiança de 95% ($OR=3.74$; IC 95% 1.23; 11.31) observa-se também que mães sem nenhum registo de carga viral tem 7 vezes a chance de ter uma criança com resultado do teste positivo comparando com uma mãe com carga viral suprimida, tendo um intervalo de confiança de 95% ($OR=7.09$; IC 1.70; 29.56).
- No abandono ao tratamento tarv, observa-se que uma mãe que tenha abandonado o tratamento tem 44 vezes a chance de ter uma criança com resultado do teste positivo comparando com uma mãe que nunca abandonou o tratamento, com um intervalo de confiança de 95% ($OR=44.35$; IC 3.20; 613.85).

Conclui-se ainda que, o presente estudo possa contribuir para melhorar as estratégias, assim como as políticas de saúde no nosso país no que concerne o acesso e a qualidade dos cuidados nas consultas das crianças em risco e pré-natais, com o objectivo de reduzir a taxa de seropositividade nas crianças expostas ao HIV/SIDA.

5.2. Recomendações

1. A realização de estudos similares de forma a identificar os factores associados à seropositividade em crianças expostas ao HIV/SIDA, de modo a trazer novos esclarecimentos assim a discussão sobre este tema;
2. Melhoramento nas estratégias assim como nas políticas de saúde no que concerne à harmonização das entidades que trabalham directamente ligado as crianças expostas ao HIV/SIDA, sobretudo por parte do parceiro privado que tem apoiado as unidades sanitárias.

5.3. Limitações

- O Centro de Saúde de Matola I foi encerrado as suas actividades de atendimento normal de pacientes, para servir de alojamento, ou seja, internamento para pacientes com COVID-19 isso no período de recolha dos dados, contribuindo assim para não recolha de dados naquela unidade sanitária razão pela qual amostra do estudo era 309, porém foi considerado 300.
- Não disponibilidade de estudos similares publicados na área de saúde à nível do distrito assim como a nível nacional.

Referências Bibliográficas

1. Batista, A.M.S. (2015). Regressão logística: uma Introdução ao modelo estatística- exemplo de aplicação ao revolving credit. Porto: Vida Económica;
2. Bertoldo, C.F. (2014). Transmissão vertical do HIV no período gestacional. Universidade de Minas Gerais: Curso de especialização em atenção básica em saúde da família. Uberaba-Brasil;
3. Carvalho, P. M. G (2015). Vulnerabilidade de Crianças expostas ao HIV/Aids por Transmissão vertical. São Paulo;
4. Cabral, C.I.S. (2013). Aplicação de Modelo de Regressão Logística num Estado de Mercado. Dissertação (Mestrado)- Universidade de Lisboa. Portugal;
5. Calvo, K. (2021). Factores Associados as Perdas de Seguimento de Crianças Expostas ao HIV: Estudo no Município de Porto Alegre. Brasil;
6. CIOB (Centro de Investigação Operacional da Beira), INS (Instituto Nacional de Saúde), UNC (Carolina Population Center), FHI360 (The Science of Improving Lives). Relatório Final: Mapeamento Programático e Prevalência de HIV entre Populações Chaves de Moçambique: Estudo PLACE 2017. Chapel Hill, NC, USA: UNC COIB e FHI360;
7. Cunha, I.V.A, Souza, B.B, Rosa, C.M.A, Iser, B.P.M, Trevisol, F.S. (2022). Factores de risco para a seroconversão de crianças expostas ao HIV no estado de santa catarina 2007-2017. Brasil;
8. Conselho Nacional de Combate ao HIV e SIDA (CNCS). 2009. Plano Estratégico Nacional de Resposta a HIV e SIDA, 2010-2014. Maputo, Moçambique: CNCS;
9. Cruz, C.C. (2012). Caracterização da morbimortalidade de crianças com HIV/SIDA em serviço de referência no sul do Brasil. São Paulo. Brasil;
10. Feracin, J.C. (2002). Atitudes e sentimentos das mulheres que vivenciaram a gravidez e a soropositividade ao vírus HIV. São Paulo: [s.n.], 2002;
11. Fávero, L e Belfiore, P (2017). Manual de Análise de Dados. 1º edição. Rio de Janeiro;
12. Figueredo, V.S.L. (2023). Transmissão vertical do HIV em uma maternidade pública em ciências médicas. Rio de Janeiro;
13. Figueira, C.V. (2006). Modelos de regressão Logística. Brasil;
14. Gonzalez, L.A. (2018). Regressão Logística e suas Aplicações. Universidade Federal do Maranhão. São Luís. Brasil;
15. Gujarati, D. N e Porter, D. C. (2011). Econometria Básica. 5ª Edição. New York. EUA

16. Hochaman, B., Nahas, F.X., Filho, R.S.O., e Ferreira, L.M. (2005). Desenhos de pesquisa: Acta Cir Bras [Serial Online].20 Suppl. 2:02-9. Disponível em URL: <http://www.scielo.br/acb>;
17. Hosmer, J. e Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley Sons Inc;
18. Hosmer, J. e Lemeshow, S. (2000). Applied Logistic Regression. 2nd ed. New York: John Wiley e Sons;
19. Incida. (2009). Inquérito Nacional de Prevalência, Riscos Comportamentais e Informação sobre o HIV e SIDA em Moçambique. Maputo: Instituto Nacional de Estatística;
20. Incida. (2021). Inquérito Nacional de Prevalência, Riscos Comportamentais e Informação sobre o HIV e SIDA em Moçambique. Maputo: Instituto Nacional de Estatística;
21. Kreitchmann, R. (2003). Transmissão perinatal do HIV-1 entre participantes do programa de controle do HIV/AIDS no sul do Brasil: um estudo de corte. Universidade Federal do Rio Grande do Sul-Dissertação de Mestrado. Brasil;
22. Lima, E. (2018). Análise de Dados Categóricos - Modelos Lineares Generalizados. Universidade Federal de Minas Gerais. Minas Gerais, Brasil;
23. Loquiha, R. A. (2015). Estatística Multivariada. Maputo: UEM;
24. Maia, A.G. (2017). Econometria: conceitos e aplicações. Cap.13.Brasil;
25. Mahaluça, F. (2016). Manual de Estatística Aplicada. Moçambique: Maputo (ISCAM);
26. Mangas, E.J.(2019). Introdução a regressão categorial aplicação a dados de escolas Angolanas, com utilização do SPSS. Brasil;
27. Maroco, J. (2007). Análise Estatística com Utilização do SPSS. 5ª edição. Lisboa. Edições Sílabo.;
28. Mesquita, C. (2014). Um modelo de regressão logística para avaliação dos programas de pós-graduação no Brasil. Rio de Janeiro, Brasil;
29. Mestre, P. (2009). Aplicação dos Modelos Lineares Generalizados às Telecomunicações Móveis: Caracterização dos clientes que desativam os seus serviços. Lisboa, Portugal;
30. Ministério de Saúde (2018). Direcção Nacional de Saúde Pública – Programa Nacional de Controle das ITS/HIV e SIDA. Moçambique;
31. Mores, M. (2007). Introdução a Análise Multivariada. 1ª Edição. Brasil;
32. Morôco, J. (2011). Análise Estatística com o SPSS statistics. 5ª Edição. Pinheiro. ReportNumber;
33. Nascimento, F.P.(2016). Classificação da pesquisa e abordagens metodológicas. Brasil;
34. Nishimoto, T.M.I., Neto, J.E., e Rozman, M.A. (2004). Transmissão materno – Infantil do vírus da imunodeficiência humana: Avaliação de medidas de controle no município de Santos. Brasil;

35. Oliveira, L.L.D., Gonçalves, A.D.C., Costa, J.S.D., e Bonilha, A.L.D.L. (2016). Factores maternos e neonatais relacionados à prematuridade. *Revista da Escola de Enfermagem da USP*, 50, 382-289;
36. Oliveira, M. F. Metodologia científica: um manual para a realização de pesquisas em Administração / Maxwell Ferreira de Oliveira. -- Catalão: UFG;
37. Onusida. (2015). *Unaid Reporto n the Global AIDS epidemic*. Ginebra;
38. Pereira, B.S., Costa, M.C.O., Amaral, M.T.R., Costa, H.S., e Silva, C.A.L., Sampaio, V.S. (2013). Factores associados à infecção pelo HIV/AIDS entre adolescentes e adultos jovens matriculados em centro de testagem e aconselhamento no estado da Bahia. *Modelo VI. Novo Horizonte*. 44.031-460. Brasil;
39. Petrie, A., & Watson, P. (2013). *Statistics for Veterinary and Animal Science*. West Sussex: A John Wiley & Sons, Ltd., Publication. Obtido em 20 de Fevereiro de 2023;
40. Salgueiro, B., Só, F., Galvão, J. e Parker, R. (2014). Associação Brasileira de AIDS-ABIA. Rio de Janeiro;
41. Salvian, M. (2016). *Multicolinearidade*. Universidade de São Paulo. Piracicba. Brasil;
42. Santos, I.P. (2016). *Introdução a análise de confiabilidade: uma aplicação ao sector de transporte*. Trabalho de licenciatura-Universidade estadual da Paraíba-PB;
43. Santos, R.C.S., e Souza, M.J.A. (2012). *Artigo de revisão de literatura: HIV na gestação*: Universidade Federal do Amapá. Brasil;
44. Silva, A., (2011). *Análise Estatística de Inquéritos online*. Rio Grande do Norte, Brasil;
45. Silva, A.N.F.(2004). *Estudo de Evolução das crianças expostas ao HIV e notificados pelo núcleo de vigilância de 2004*. São paulo;
46. Silva, J.A., Dourado, I., Brito, A.M., e Silva, C.A.L. (2015). Factores associados à não adesão ao antirretrovirais em adultos com AIDS nos seis primeiros meses da terapia em salvador. *Cad. Saúde Pública*. Rio de Janeiro. 31(6):1188-1198;
47. Saidl et al.(2005). *Crianças e adolescentes vivendo com HIV e suas familias aspectos psicossociais*. Brasil;
48. Silveira, M.B.G. (2021). *Aplicação da regressão logística na análise dos factores de risco associados a hipertensão arterial*. Universidade Estadual da Paraíba- Curso de Bacharelado em Estatística. Brasil;
49. Siqueiro, P.G., Duarte, G. M., e Souza, M.V. (2020). *Análise hierarquizada dos determinantes da transmissão vertical do HIV: um estudo de caso-controle*. Brasil;
50. Turkman, M.A.A., e Silva, G.L. (2000). *Modelos Lineares Generalizados da Teórica à Prática*. In. *Viii Congresso Anual da Sociedade Portuguesa de Estatística*. Lisboa;

51. UNAIDS/WHO. 2009. AIDS epidemic update. Genebra: UNAIDS/WHO;
52. Vaz, M.J.R., Barros, S.M.O., e Rodrigues, S.T.C. (2011). Transmissão vertical do HIV em população atendida no serviço de referência. São Paulo. Brasil;
53. Velho, S.D.F. (2015). Modelos de Sobrevivência para Estudo do Tempo até a ocorrência de Excesso de Peso em Indivíduos Adultos Submetidos à Transplante Alogénico de Células Progenitoras Hematopoiéticas. Dissertação (Mestrado) – Universidade de Lisboa;
54. World Health Organization [WHO]. (2023, July 13). *HIV and AIDS*. (World Health Organization) Retrieved September 19, 2023, from World Health Organization: https://www.who.int/news-room/fact-sheets/detail/hiv-aids?gclid=CjwKCAjwu4WoBhBkEiwAojNdXkFkZixWaSdDdLLJsFxpkgfn9QaenlJmd0ZorSWovYYe4ju0DU3OVxoCZJsQAvD_BwE.

Anexos

Código

```
---  
title: "Alfredo"  
output: html_document  
date: "2024-03-14"  
---  
  
`` {r setup, include=FALSE}  
knitr::opts_chunk$set(echo = TRUE)  
``  
  
## R Markdown  
  
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF,  
and MS Word documents. For more details on using R Markdown see  
<http://rmarkdown.rstudio.com>.
```

*****COMANDOS USADOS NA ANÁLISE DE DADOS****

*****Instalação e carregamento pacotes necessários para a Base_de_dados3****

```
`` {r, warning=FALSE, message=FALSE}  
if(!require(pacman)) install.packages("pacman")  
library(pacman)  
if(!require(stargazer)) install.packages("stargazer")  
library(stargazer)  
if(!require(dplyr)) install.packages("dplyr")  
library(dplyr)  
if(!require(RVAideMemoire)) install.packages("RVAideMemoire")  
library(RVAideMemoire)  
if(!require(car)) install.packages("car")  
library(car)  
if(!require(psych)) install.packages("psych")  
library(psych)  
if(!require(rstatix)) install.packages("rstatix")  
library(rstatix)  
if(!require(DescTools)) install.packages("DescTools")  
library(DescTools)  
if(!require(ggplot2)) install.packages("ggplot2")  
library(ggplot2)  
require(haven)  
library(haven)  
if(!require(rminer)) install.packages("rminer")  
library(rminer)  
if(!require(plotly)) install.packages("plotly")
```



```

library(plotly)
if(!require(caret)) install.packages("caret")
library(caret)
if(!require(cluster)) install.packages("cluster")
library(cluster)
if(!require(survival)) install.packages("survival")
library(survival)
if(!require(sampling)) install.packages("sampling")
library(sampling)
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)
if(!require(pROC)) install.packages("pROC")
library(pROC)
if(!require(ResourceSelection)) install.packages("pROC")
library(ResourceSelection)
if(!require(BatchGetSymbols)) install.packages("BatchGetSymbols")
library(BatchGetSymbols)
if(!require(openxlsx)) install.packages("openxlsx")
library(openxlsx)
if(!require(knitr)) install.packages("knitr")
library(knitr)
if(!require(rmarkdown)) install.packages("rmarkdown")
library(rmarkdown)
if(!require(emmeans)) install.packages("emmeans")
library(emmeans)
if(!require(kableExtra)) install.packages("kableExtra")
library(kableExtra)
if(!require(htmltools)) install.packages("htmltools")
library(htmltools)
...

###Carregar a Base de dados "Base_de_dados2"
````{r, warning=FALSE, message=FALSE}
setwd("C:/Users/ACER/Documents/baule.xlsx")
library(readxl)
Base_de_dados2 <- read_excel("Base_de_dados2")
View(Base_de_dados2)
rm(GCtorture)

###Converter as variáveis da base de dados "Base_de_dados2" para variáveis factor
````{r, warning=FALSE, message=FALSE}
Base_de_dados2$s_xo = as.factor(Base_de_dados2$s_xo)
Base_de_dados2$ti_parto = as.factor(Base_de_dados2$ti_parto)
Base_de_dados2$alei_mto = as.factor(Base_de_dados2$alei_mto)
Base_de_dados2$ptv_Cria = as.factor(Base_de_dados2$ptv_Cria)
Base_de_dados2$a_trat_mae = as.factor(Base_de_dados2$a_trat_mae)
Base_de_dados2$a_famly = as.factor(Base_de_dados2$a_famly)
Base_de_dados2$c_sa = as.factor(Base_de_dados2$c_sa)
Base_de_dados2$f_aga = as.factor(Base_de_dados2$f_aga)
Base_de_dados2$f_quiter = as.factor(Base_de_dados2$f_quiter)
Base_de_dados2$o_pcao = as.factor(Base_de_dados2$o_pcao)
Base_de_dados2$a_metcao = as.factor(Base_de_dados2$a_metcao)
Base_de_dados2$m_parcer = as.factor(Base_de_dados2$m_parcer)

```

```

Base_de_dados2$r_giao = as.factor(Base_de_dados2$r_giao)
Base_de_dados2$m_dimnto = as.factor(Base_de_dados2$m_dimnto)
Base_de_dados2$c_viral = as.factor(Base_de_dados2$c_viral)
Base_de_dados2$a_bandno = as.factor(Base_de_dados2$a_bandno)
Base_de_dados2$i_ncio = as.factor(Base_de_dados2$i_ncio)
Base_de_dados2$r_test = as.factor(Base_de_dados2$r_test)
...

```

*****Identificar as categorias de referencia das variaveis na base de dados "Base_de_dados2"*****

```

```{r, warning=FALSE, message=FALSE}
Base_de_dados2$s_xo = relevel(Base_de_dados2$s_xo, ref = "Masculino")
Base_de_dados2$ti_parto = relevel(Base_de_dados2$ti_parto, ref = "Cesariana")
Base_de_dados2$alei_mto = relevel(Base_de_dados2$alei_mto, ref = "Artificial")
Base_de_dados2$ptv_Cria = relevel(Base_de_dados2$ptv_Cria, ref = "NVP + AZT")
Base_de_dados2$a_trat_mae = relevel(Base_de_dados2$a_trat_mae, ref = "< 3 anos")
Base_de_dados2$a_famly = relevel(Base_de_dados2$a_famly, ref = "<=3 AF")
Base_de_dados2$c_sa = relevel(Base_de_dados2$c_sa, ref = "Canico")
Base_de_dados2$f_aga = relevel(Base_de_dados2$f_aga, ref = "Poco")
Base_de_dados2$f_quiter = relevel(Base_de_dados2$f_quiter, ref = "Sim")
Base_de_dados2$o_pcao = relevel(Base_de_dados2$o_pcao, ref = "Emprego")
Base_de_dados2$a_metcao = relevel(Base_de_dados2$a_metcao, ref = "Não")
Base_de_dados2$m_parcer = relevel(Base_de_dados2$m_parcer, ref = "Não")
Base_de_dados2$r_giao = relevel(Base_de_dados2$r_giao, ref = "Não")
Base_de_dados2$m_dimnto = relevel(Base_de_dados2$m_dimnto, ref = "sim toma")
Base_de_dados2$c_viral = relevel(Base_de_dados2$c_viral, ref = "Suprimida")
Base_de_dados2$a_bandno = relevel(Base_de_dados2$a_bandno, ref = "Não")
Base_de_dados2$i_ncio = relevel(Base_de_dados2$i_ncio, ref = "Não")
Base_de_dados2$r_test = relevel(Base_de_dados2$r_test, ref = "Negativo")
...

```

**\*\*\*Verificar e confirmar as categorias de referencia identificadas das variáveis na base de dados "Base\_de\_dados2"\*\*\***

```

```{r, warning=FALSE, message=FALSE}
levels(Base_de_dados2$s_xo)
levels(Base_de_dados2$ti_parto)
levels(Base_de_dados2$alei_mto)
levels(Base_de_dados2$ptv_Cria)
levels(Base_de_dados2$a_trat_mae)
levels(Base_de_dados2$a_famly)
levels(Base_de_dados2$c_sa)
levels(Base_de_dados2$f_aga)
levels(Base_de_dados2$f_quiter)
levels(Base_de_dados2$o_pcao)
levels(Base_de_dados2$a_metcao)
levels(Base_de_dados2$m_parcer)
levels(Base_de_dados2$r_giao)
levels(Base_de_dados2$m_dimnto)
levels(Base_de_dados2$c_viral)
levels(Base_de_dados2$a_bandno)
levels(Base_de_dados2$i_ncio)
levels(Base_de_dados2$r_test)
...

```

*****Realizar teste de associação "teste Qui_Quadrado" e/ou "teste Exacto de Fisher"*****

###Para teste Qui_Quadrado****

```
`` {r, warning=FALSE, message=FALSE}
chisq.test(Base_de_dados2$r_test, Base_de_dados2$s_xo)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$ti_parto)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$alei_mto)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$ptv_Cria)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$a_trat_mae)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$a_famly)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$c_sa)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$f_aga)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$f_quiter)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$o_pcao)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$a_metcao)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$m_parcer)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$r_giao)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$m_dimnto)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$c_viral)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$a_bandno)
chisq.test(Base_de_dados2$r_test, Base_de_dados2$i_ncio)
``
```

###Para teste Exacto de Fisher****

```
`` {r, warning=FALSE, message=FALSE}
fisher.test(Base_de_dados2$r_test, Base_de_dados2$s_xo)

fisher.test(Base_de_dados2$s_xo, Base_de_dados2$alei_mto)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$ptv_Cria)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$a_trat_mae)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$a_famly)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$c_sa)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$f_aga)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$f_quiter)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$o_pcao)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$a_metcao)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$m_parcer)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$r_giao)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$m_dimnto)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$c_viral)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$a_bandno)
fisher.test(Base_de_dados2$s_xo, Base_de_dados2$i_ncio)
#fisher.test(Base_de_dados2$s_xo, Base_de_dados2$r_test)
``
```

*****Construção de regressão logística****

###Verificados pressupostos prévios para criação da regressão logística onde:****

- + 1. A variável dependente no presente Base_de_dados2 é dicotómica e mutuamente exclusiva
- + 2. Existe independência das observações (sem medidas repetidas)

*****Construção de modelos****

```

###*Contruindo o modelo 1 = "modelo1" contendo todas as 18 variáveis consideradas no
Base_de_dados2**
```{r, warning=FALSE, message=FALSE}
modelo1 = glm(r_test ~ s_xo + ti_parto + alei_mto + ptv_Cria
 + a_trat_mae + a_famly + c_sa + f_aga + f_quiter
 + o_pcao + a_metcao + m_parcer +
 r_giao + m_dimnto + c_viral + a_bandno +
 i_ncio + r_test ,
 family = binomial(link = 'logit'), data = Base_de_dados2)
...

###*Base_de_dados3 de “modelo1” = Overall effects = efeito geral do “modelo1”*
```{r, warning=FALSE, message=FALSE}
Anova(modelo1,type = 'II', test = 'F')
...

###*efeitos especificos do "modelo1"*
```{r, warning=FALSE, message=FALSE}
summary(modelo1)
...

###*Contruindo o modelo 2 = “modelo2” contendo todas as variáveis do Base_de_dados2 com
efeito significativo na Base_de_dados3 bivariada**
```{r, warning=FALSE, message=FALSE}
modelo2 = glm(r_test ~ ptv_Cria + a_metcao + r_giao + m_dimnto + c_viral
  + a_bandno + i_ncio ,
  family = binomial(link = 'logit'), data = Base_de_dados2)
...

###*Base_de_dados3 de “modelo2” = Overall effects = efeito geral do “modelo2”*
```{r, warning=FALSE, message=FALSE}
Anova(modelo2, type = "II", test = "F")
...

###*efeitos especificos do "modelo2"*
```{r, warning=FALSE, message=FALSE}
summary(modelo2)
...

###*Comparar a qualidade dos modelos 1, 2**
###*Pseudo R-quadrado (explica a % de variância do modelo)*
```{r, warning=FALSE, message=FALSE}
PseudoR2(modelo1, which = "Nagelkerke")
PseudoR2(modelo2, which = "Nagelkerke")
...

###*Comparação com AIC e BIC (se a diferença for pelo menos 10 o melhor modelo é o que
apresentar menor valor)**
###*AIC: Akaike’s Information Criteria (AIC) ou Critério de Informação de Akaike*
```{r, warning=FALSE, message=FALSE}
AIC(modelo1, modelo2)
...

###*BIC ou SBC: Schwarz Bayesian Information Criteria ou Critério de Informação Bayesiano
de Schwarz*
```{r, warning=FALSE, message=FALSE}

```

```
BIC(modelo1, modelo2) #melhor modelo #2
```

```
####comparacao atraves do Qui_Quadrado - #se p > 5% os modelos nao sao estatisticamente significativos*
```

```
` `{r, warning=FALSE, message=FALSE}
BIC(modelo1, modelo2)
"
```

```
####Verificar os pressupostos do modelo aprovado "modelo2"*
####3. Ausencia de outliers*
```

```
` `{r}
par(mfrow=c(2,2))
plot(modelo2)
```  
  
ou  
` `{r, warning=FALSE, message=FALSE}  
par(mfrow=c(1,1))  
plot(modelo2, wch =5)  
```
```

```
ou pedir resíduos padronizados
```

```
` `{r, warning=FALSE, message=FALSE}
summary(rstandard(modelo2))
```
```

```
####4. Ausencia de Multicolinearidade se r>0.9 (ou 0.8) ou VIF > 10*
```

```
` `{r, warning=FALSE, message=FALSE}  
library(readxl)  
Base_de_dados3 <- read_excel("Base_de_dados3.xlsx")
```

```
Base_de_dados3$r_test = as.factor(Base_de_dados3$r_test)  
Base_de_dados3$ptv_Cria = as.factor(Base_de_dados3$ptv_Cria)  
Base_de_dados3$a_metcao = as.factor(Base_de_dados3$a_metcao)  
Base_de_dados3$r_giao = as.factor(Base_de_dados3$r_giao )  
Base_de_dados3$m_dimnto = as.factor(Base_de_dados3$m_dimnto)  
Base_de_dados3$c_viral = as.factor(Base_de_dados3$c_viral)  
Base_de_dados3$a_bandno = as.factor(Base_de_dados3$a_bandno)  
Base_de_dados3$i_ncio = as.factor(Base_de_dados3$i_ncio)
```

```
glimpse(Base_de_dados3)
```

```
Base_de_dados3$r_test = relevel(Base_de_dados3$r_test, ref = "Negativo")  
Base_de_dados3$ptv_Cria = relevel(Base_de_dados3$ptv_Cria, ref = "NVP + AZT")  
Base_de_dados3$a_metcao = relevel(Base_de_dados3$a_metcao, ref = "Não")  
Base_de_dados3$r_giao = relevel(Base_de_dados3$r_giao, ref = "Não")  
Base_de_dados3$m_dimnto = relevel(Base_de_dados3$m_dimnto, ref = "sim toma")  
Base_de_dados3$c_viral = relevel(Base_de_dados3$c_viral, ref = "Suprimida")  
Base_de_dados3$a_bandno = relevel(Base_de_dados3$a_bandno, ref = "Não")  
Base_de_dados3$i_ncio = relevel(Base_de_dados3$i_ncio, ref = "Não")
```

```
levels(Base_de_dados3$r_test)  
levels(Base_de_dados3$ptv_Cria)
```

```

levels(Base_de_dados3$a_metcao)
levels(Base_de_dados3$r_giao)
levels(Base_de_dados3$m_dimnto)
levels(Base_de_dados3$c_viral)
levels(Base_de_dados3$a_bandno)
levels(Base_de_dados3$i_ncio)
#ou
`` {r, warning=FALSE, message=FALSE}
pairs.panels(Base_de_dados3)
``

ou VIF nao deve ser > 10
`` {r, warning=FALSE, message=FALSE}
modelo2.1 = glm(r_test ~ ptv_Cria + a_metcao + r_giao + m_dimnto + c_viral
+ a_bandno + i_ncio ,
family = binomial(link = 'logit'), data = Base_de_dados3)

summary(modelo2.1)
vif(modelo2.1)

###Base_de_dados3 de **
##**Base_de_dados3 de = Overall effects = efeito geral do **
`` {r, warning=FALSE, message=FALSE}
Anova(modelo2.1,type = 'II', test = 'F')

#Teste de razao de verosimilhanca
Anova(modelo2.1, test = "LR")

###**Efeitos específicos**
`` {r, warning=FALSE, message=FALSE}
summary(modelo2.1)
``

###**OObter Odds Ratio (OR) com IC 95% para **
`` {r, warning=FALSE, message=FALSE}
exp(coef(modelo2.1))
confint(modelo2.1)
exp(cbind(OR = coef(modelo2.1), confint.default(modelo2.1)))

###**Matriz de Confusão ou tabela de classificação do modelo (quanto o modelo acertou)**
`` {r, warning=FALSE, message=FALSE}
ClassLog(modelo2.1, Base_de_dados3$r_test)

###**Recriando o modelo "modelo2**
`` {r, warning=FALSE, message=FALSE}
modelo2.1 = glm(r_test ~ ptv_Cria + a_metcao + r_giao + m_dimnto + c_viral
+ a_bandno + i_ncio ,
family = binomial(link = 'logit'), data = Base_de_dados3)

summary(modelo2.1)

```

```

...
###**Criando campo de predição para cada idade dos indivíduos**
```{r, warning=FALSE, message=FALSE}
Base_de_dados3$PRED=predict(modelo2.1, newdata=Base_de_dados3, type="response")
...

###**Plotando a probabilidade predita pelo **
```{r, warning=FALSE, message=FALSE}
ggplot(Base_de_dados3, aes(x=Base_de_dados3$c_viral, y=Base_de_dados3$PRED)) +
  geom_point()
...

###**Cálculo precisão, especificidade e sensibilidade**
```{r, warning=FALSE, message=FALSE}
Base_de_dados3$pdata <- as.factor(
 ifelse(
 predict(modelo2.1,
 newdata = Base_de_dados3,
 type = "response")
 >0.5,"Positivo","Negativo"))
confusionMatrix(Base_de_dados3$pdata, Base_de_dados3$r_test, positive="Positivo")
...

```{r, warning=FALSE, message=FALSE}
confusionMatrix(Base_de_dados3$pdata, Base_de_dados3$r_test, positive="Positivo")
...

#####Curva ROC**
```{r, warning=FALSE, message=FALSE}
roc1 = plot.roc(Base_de_dados3$r_test,fitted(modelo2.1))
...

```{r, warning=FALSE, message=FALSE}
plot(roc1,
  print.auc=TRUE,
  auc.polygon=TRUE,
  grid=c(0.1,0.2),
  grid.col=c("green","red"),
  max.auc.polygon=TRUE,
  auc.polygon.col="lightgreen",
  print.thres=TRUE)
###**O teste Hosmer e Lemeshow**
```{r, warning=FALSE, message=FALSE}
hlmodelo2.1=hoslem.test(Base_de_dados3$r_test,fitted(modelo2.1),g=10)
hlmodelo2.1

...

###**Exportar os dados para o Excel**
```{r, warning=FALSE, message=FALSE}
getwd()
setwd("C:/Users/ACER/Downloads")
aquivodesaida = 'probabilidade.xlsx'
write.xlsx(Base_de_dados3,aquivodesaida)
...

```

1. Estimaco do modelo1 de regresso logstica com todas as variveis

Coefficients:

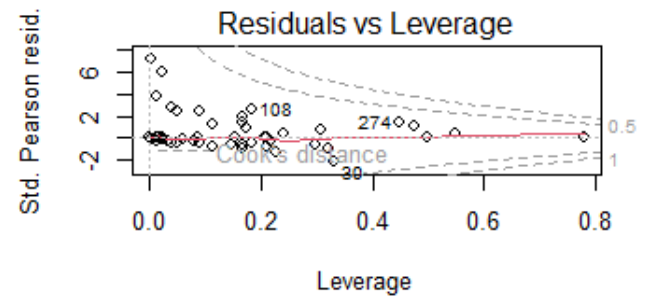
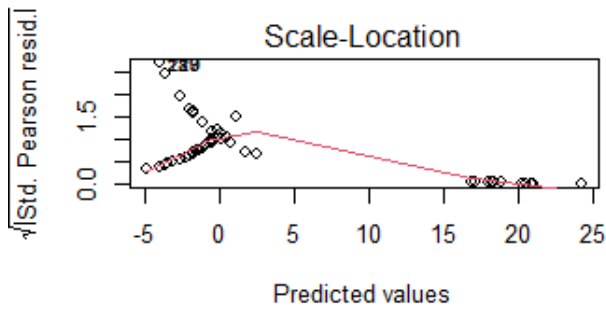
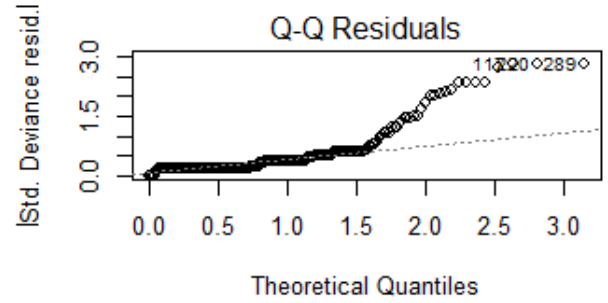
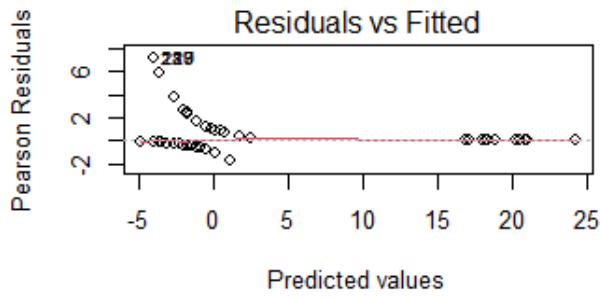
	Estimate	Std. Error	z	value	Pr(> z)
(Intercept)	-34.20209	6765.18749	-0.005	0.995966	
s_xoFeminino	0.01678	0.60524	0.028	0.977877	
ti_partoNormal	14.81867	1795.25248	0.008	0.993414	
alei_mtoMaterno	12.73900	6522.63878	0.002	0.998442	
alei_mtoMisto	14.56647	6522.63887	0.002	0.998218	
ptv_CriaAZT	0.52186	1.10285	0.473	0.636078	
ptv_CriaNenhum	2.62918	1.58979	1.654	0.098172	.
ptv_CriaNVP	1.06345	0.99826	1.065	0.286738	
a_trat_mae> 6 anos	0.29844	0.96004	0.311	0.755906	
a_trat_mae3 - 6 anos	-0.17822	0.68236	-0.261	0.793948	
a_family>=7 AF	-0.30027	1.14247	-0.263	0.792687	
a_family4 - 6 AF	0.72919	1.09320	0.667	0.504753	
c_saBloco	-0.64309	3.46460	-0.186	0.852745	
f_agaAgua + de 50 metros	5.46983	4.25523	1.285	0.198640	
f_agaAgua Canalizada	1.28818	3.71202	0.347	0.728569	
f_quiterNo	0.84712	1.10151	0.769	0.441862	
o_pcaoAgricultura de Subsistencia	-0.30507	1.41395	-0.216	0.829177	
o_pcaoAuto - Emprego	-0.14296	0.96533	-0.148	0.882266	
o_pcaoNenhum	1.46660	0.88594	1.655	0.097838	.
a_metcaosas vezes	2.12022	0.96473	2.198	0.027968	*
a_metcaosim	22.21218	2404.31341	0.009	0.992629	
m_parcersim	2.06030	0.75202	2.740	0.006150	**
r_giaosim	21.08227	2229.52518	0.009	0.992455	
m_dimntoFalha na toma	2.99914	0.90001	3.332	0.000861	***
m_dimntoNo toma	-0.61489	3.85658	-0.159	0.873322	
c_viralNao Suprimida	1.40194	0.64620	2.170	0.030044	*
c_viralNenhuma	3.07540	0.90464	3.400	0.000675	***
a_bandnosim	4.93042	1.63710	3.012	0.002598	**
i_nciosim	3.13518	1.34956	2.323	0.020173	*

2. Estimaco do modelo2 de regresso logstica com apenas variveis significativas

Coefficients:

	Estimate	Std. Error	z	value	Pr(> z)
(Intercept)	-3.9570	0.4904	-8.070	7.06e-16	***
ptv_CriaAZT	0.3857	0.9197	0.419	0.674954	
ptv_CriaNenhum	2.4982	1.4227	1.756	0.079083	.
ptv_CriaNVP	0.9205	0.8476	1.086	0.277472	
a_metcaosas vezes	2.1921	0.8335	2.630	0.008537	**
a_metcaosim	20.8350	2475.6938	0.008	0.993285	
r_giaosim	20.6348	2245.7529	0.009	0.992669	
m_dimntoFalha na toma	2.2019	0.6578	3.348	0.000815	***
m_dimntoNo toma	-0.8679	2.7374	-0.317	0.751210	
c_viralNao Suprimida	1.3204	0.5641	2.341	0.019241	*
c_viralNenhuma	1.9598	0.7280	2.692	0.007098	**
a_bandnosim	3.7922	1.3406	2.829	0.004673	**
i_nciosim	2.1966	1.1989	1.832	0.066918	.

3. Análise Residual



Resíduos Padronizados

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.0660	-0.3743	-0.1951	-0.1435	-0.1951	2.8262



UNIVERSIDADE EDUARDO MONDLANE

Faculdade de Ciências
Departamento de Matemática e Informática
Curso de Estatística

QUESTIONÁRIO

Com o objectivo de avaliar os factores associados a seropositividade em crianças Expostas ao vírus do HIV/SIDA no Distrito da Matola, este questionário pretende colher informações nas Unidades Sanitárias da Matola em relação aos diversos serviços prestados pela Saúde Materno Infantil (SMI) concretamente na Consulta da Criança em Risco (CCR). O questionário é de natureza confidencial, e sobretudo o anonimato será respeitado, visto que os dados recolhidos serão usados apenas para fins de pesquisa.

Variáveis de Estudo

1. Sexo:

- Masculino
 Femenino

2. Tipo de Parto:

- Normal
 Cesariana

3. Aleitamento:

- Materno
 Artificial
 Misto

4. PTV Criança:

- NVP
 AZT
 NVP + AZT

5. Anos em Tarv da Mãe:

- ≤ 3 anos
 $> 3 - 6$ anos
 > 6 anos

6. Última Carga Viral da Mãe:

- Nenhuma
 Suprimida
 Não Suprimida

7. Agregado Familiar:

- ≤ 3 AF
 $4 - 6$ AF
 ≥ 7 AF

8. Casa:

- Bloco
 Caniço

9. Fonte de Água:

- Poço
 Água Canalizada
 Fonte de água mais de 50 metros

10. Rede Mosquiteira:

- Sim
 Não

11. Ocupação da mãe:

- Agricultura de Subsistência
 Auto-Emprego
 Emprego
 Nenhum

11. Multiplos parceiros:

- Sim
 Não

12. Falta de Alimentação da mãe HIV+

- As vezes
 Não
 Sim

13. Crenças religiosas para não adesão ao Tarv

- Sim
 Não

14. Uso de medicamentos Antirretrovirais pela mãe HIV+

- Falha na Toma
 Não Toma
 Sim Toma

15. Abandono da mãe ao Tarv

- Sim
 Não

16. Início Tarv Tardamente da mãe HIV+ durante a gestação

- Sim Não



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Ao
Serviço Distrital de Saúde, Mulher e Acção Social da Matola
Unidades Sanitárias do Distrito da Matola
Maputo

Credencial

Como é do domínio público, a **Universidade Eduardo Mondlane - UEM**, ministra diversos cursos formando técnicos superiores capacitados para responder à demanda de várias solicitações que o mercado oferece. Neste contexto, Credencia - se para efeitos de recolha dados, o Sr. **Alfredo Baúle Júnior**, estudante do 4º ano do Curso de Licenciatura em Estatística na Unidade onde V.Excia é digno dirigente.

Sem mais, subscrevemo-nos com protestos de elevada estima e consideração, na expectativa de que este assunto venha merecer o melhor acolhimento.

Cordiais Saudações.

Maputo, 18 de Novembro de 2022
O Director do Curso de Estatística

Doutor Miranda Albino Martins Muualalo





REPÚBLICA DE MOÇAMBIQUE
PROVÍNCIA DE MAPUTO
GOVERNO DO DISTRITO DA MATOLA
SERVIÇO DISTRITAL DE SAÚDE, MULHER E ACÇÃO SOCIAL

Ao:

Centro de Saúde de Ndlavela
Matola

199)
Ref. n° / RH/024.1/2022

Data: 30/11/2022

Assunto: Comunicação

Para os devidos efeitos, comunica se a Unidade Sanitária que o funcionário e estudante **Alfredo Baúle Júnior**-Técnica de Saúde irá efectuar a recolha de dados em todos Centros de Saúde, no período de 05 de Dezembro a 23 de Dezembro de 2022.

Cordiais saudações

CLN/cln

= Cláudia Julieta Ernesto Ndongue José =

Técnico Superior NI/



AV:5 de Fevereiro n° 812/Serviço de Saúde, Mulher e Acção Social da Matola

Visão
Amo Muhalo
Emf - geral
13.12.2022



REPÚBLICA DE MOÇAMBIQUE

PROVÍNCIA DE MAPUTO
GOVERNO DO DISTRITO DA MATOLA
SERVIÇO DISTRITAL DE SAÚDE, MULHER E ACÇÃO SOCIAL DA MATOLA

CREDÊNCIAL

Por despacho da Senhora Directora do Serviço Distrital de Saúde, Mulher e Acção Social da Matola, esta devidamente credenciado o Senhor **Alfredo Baúle Júnior**, estudante do 4º ano do curso de Licenciatura em Estatística na Universidade Eduardo Mondlane, Faculdade de Ciências, a fim de recolha de dados, em todos os Centro de Saúde, a partir do dia 05 de Dezembro a 23 de Dezembro de 2022.

Matola, aos 30 de Novembro de 2022

Enrole no OVS
e apresente
na U.S. no dia
19.12.22, no site no mesmo
dia 19.12.22
Monte

A Chefe da Repartição RH

Cláudia Julieta F. Longue
/Técnica Superior N1/

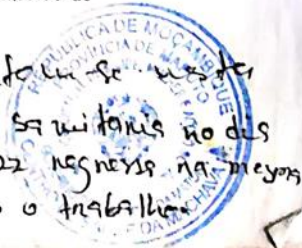
Apresentou-se
na U.S.
16-12-2022
Semora

Apresentou-se
na U.S. e volta
a credenciado
14.12.2022



Serviço Distrital de Saúde, Mulher e Acção Social da Matola - Matola 700 / Maputo Provincial, Av. 5 de Fevereiro nº-812

Apresentou-se na
unidade sanitária no dia
16.12.2022 regressa na mesma
data após o trabalho



Apresentou-se na Unidade Sanitária -
Centro de Saúde Ntavele,
Est. Luísa
16/12/2022

Apresentou-se nesta unidade
Sanitária do C. S. M-Gare
fm de levantadas.
20/12/2022

REPUBLICA DE MOÇAMBIQUE
SERVIÇO DISTRITAL DE SAÚDE, MULHER E ADOLESCÊNCIA
Centro de Saúde de Tsalala
Entrada n.º 025-3
Data 20/12/22
Assinatura: [Signature]

Apresentou-se no
C.S. Tsalala de 20
de 20/12/22

Apresentou-se no
C.S. de 21/12/22
Registo de exames de
[Signature]

REPUBLICA DE MOÇAMBIQUE
SDSMAS - MATICA
Entrada n.º 31-025-3
BOS... 20/12/22
BOS... 20/12/22



Apresentou-se
no C.S. de 21-12-22
legitimada por [Signature]



Apresentou-se na
Unidade Sanitária
de 21/12/22
[Signature]

Apresentou-se no
C.S. de 27/12/22
[Signature]



REPÚBLICA DE MOÇAMBIQUE

PROVÍNCIA DE MAPUTO
GOVERNO DO DISTRITO DA MATOLA
SERVIÇO DISTRITAL DE SAÚDE, MULHER E ACÇÃO SOCIAL DA MATOLA

CREDÊNCIAL

Por despacho da Senhora Directora do Serviço Distrital de Saúde, Mulher e Acção Social da Matola, esta devidamente credenciado o Senhor **Alfredo Baúle Júnior**, estudante do 4º ano do curso de Licenciatura em Estatística na Universidade Eduardo Mondlane, Faculdade de Ciências, a fim de recolha de dados, em todos os Centro de Saúde, nos dias 28 de Dezembro a 30 de Dezembro de 2022.

Centro de Saúde da Liberdade
Entrada n.º 100.025-5
Data 20/12/22
Ass. Cláudia

Matola, 23 de Dezembro de 2022

Alfredo Baúle Júnior
O Chefe da RAM
Salesio Orlick Néves de Andrade
/Médico de Clínica Geral/



REPÚBLICA DE MOÇAMBIQUE
SERVIÇO DISTRITAL DE SAÚDE, MULHER E ACÇÃO SOCIAL
CENTRO DE SAÚDE DE NITÓRE
Entrada N.º 100.025-5
Data 20/12/22
Assinatura Cláudia

Recebido
20/12/22
19

REPÚBLICA DE MOÇAMBIQUE
SERVIÇO DISTRITAL DE SAÚDE, MULHER E ACÇÃO SOCIAL
Centro de Saúde de Budede
Entrada N.º 100.025-5
Data 29/12/22
Assinatura Salesio Orlick Néves de Andrade